SYM-AM-22-029

Proceedings of the Nineteenth Annual Acquisition Research Symposium

Thursday, May 12, 2022 Sessions
Volume II

Acquisition Research: Creating Synergy for Informed Change

May 11–12, 2022

Published: April 29, 2022

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School

The research presented in this report was supported by the Acquisition Research Program at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School

Table of Contents

SYM-AM-22-029

Proceedings
of the
Nineteenth Annual Acquisition Research Symposium

Thursday, May 12, 2021 Sessions
Volume II

Acquisition Research:
Creating Synergy for Informed Change

May 11-12, 2022

Published: April 29, 2022

THIS PAGE INTENTIONALLY LEFT BLANK

WELCOME: DAVID H. LEWIS, VICE ADMIRAL, U.S. NAVY (RET), ACQUISITION CHAIR, ACQUISITION RESEARCH PROGRAM

Vice Admiral David H. Lewis, USN (Ret.) took the helm as the Naval Postgraduate School Chair of Acquisition. As chair, he will lead the Acquisition Research Program (ARP) in the Graduate School of Defense Management and connect NPS with leaders and policymakers in the acquisition community.

Lewis graduated from NPS in 1988 with a Master of Science in Computer Science, and we're pleased to welcome him back to campus in this leadership role. Lewis is replacing the founding Chair of Acquisition, Rear Admiral, USN (Ret.) Jim Greene, who retired this June.

Most recently, Lewis served as Director of the Defense Contract Management Agency, managing over $7 trillion in defense contracts. In this role, he oversaw the agency's efforts to ensure that supplies and services contracted for by the Department of Defense are delivered on time and in line with contract performance requirements.

During his career at sea, Lewis served as a communications officer, fire control and missile battery officer, and combat systems officer aboard destroyers and guided-missile cruisers.

Upon selection to flag rank in 2009, Lewis served as Vice Commander, Naval Sea Systems Command and then served four years as Program Executive Officer, Ships, where he directed the delivery of 18 ships and procurement of another 51 ships. From 2014-2017 he served as Commander, Space and Naval Warfare Systems Command where he led a global workforce of 10,300 civilian and military personnel who design, develop and deploy advanced communications and information capabilities.

Lewis's extensive experience in shipbuilding has given him a unique understanding of the full acquisition lifecycle. He has delivered ships as a program manager and program executive officer, then later sustained and modernized them as a fleet engineer and systems commander. He will bring valuable perspective to NPS students and faculty, as well as the broader acquisition innovation community working to get superior capabilities into the hands of our warfighters.

Lewis's expertise in product delivery will amplify ARP's ability to execute its mission of delivering the real-time information and analytical capabilities needed by today's acquisition professionals and policymakers. Adding VADM Lewis to the team also demonstrates NPS's continued commitment to providing world-class defense-focused education and research.

KEYNOTE SPEAKER: MR. TOMMY ROSS, CHIEF OF STAFF OFFICE OF THE SECRETARY OF THE NAVY

Tommy Ross serves as Chief of Staff to the Secretary of the Navy, the Honorable Carlos Del Toro. He is responsible for providing counsel and advice to the Secretary on all matters concerning the Department of the Navy.

Previously, he was Senior Director for Policy at BSA | The Software Alliance, an international trade association representing the world's leading software and data service companies. Responsible for policies relating to cyber security, 5G, and law enforcement access to digital evidence, he led development of BSA's Framework for Secure Software, a first-of-its-kind risk-based framework for evaluating security of software products and services.

Prior to joining BSA, he served as the first Deputy Assistant Secretary of Defense for Security Cooperation, responsible for prioritizing Department of Defense bilateral and multilateral security cooperation activities and aligning security cooperation resources to the defense strategy. Under his leadership, the Department undertook historic reforms to its security cooperation enterprise, including a consolidation and reform of the Department's legislative authorities, the establishment of the first-ever security cooperation Assessment, Monitoring, and Evaluation (AM&E) framework, expansion and modernization of the Department's institutional capacity-building efforts, and the initiation and implementation of the Southeast Asia Maritime Security Initiative.

From 2009 to 2014, he served as Senior Intelligence and Defense Advisor to Senator Harry Reid of Nevada, the Senate Majority Leader. In this capacity, he advised Senator Reid on national security matters, developed and implemented legislative strategies for all defense and intelligence legislative and policy initiatives in the Senate, and was responsible for all compartmented "Gang of 8" intelligence matters for the Majority Leader.

From 2005 to 2009, he served on the staff of Rep. David Price of North Carolina, first as Legislative Assistant and, subsequently as Legislative Director. He also staffed Representative Price's work as Chairman of the House Democracy Partnership, a congressional commission working to strengthen institutional capabilities of legislatures in developing democracies. Previously, he served as a policy analyst for the Senate Democratic Policy Committee, where he supported the Democratic Caucus on national security policy and communications. He began his career as a research assistant for Senate Majority Leader Tom Daschle.

Tommy has served as Chairman of the Board of Directors for Mine Action Group America and as a Senior Associate at the Center for Strategic and International Studies. He is a graduate of Davidson College in North Carolina, and holds an M.A. in Theology and Ethics from Union Theological Seminary in New York. He has completed the Air Force's Air Command and Staff College as well as a Certificate in Africa Intelligence Studies at the National Intelligence University. Tommy and his wife Lindsay are the proud parents of daughter Ruby and son Sebastian.

PANEL 14. ACQUISITION MODERNIZATION

Thursday, May 12, 2022	
8:15 a.m. – 9:30 am.	**Chair: Brigadier General Mike Sloane, USA (ret.),** Dean, Defense Systems Management College, Defense Acquisition University **Panelists:** **Ronald R. Richardson, Jr.,** Director, U.S. Army Acquisition Support Center and Director, Acquisition Career Management (DACM) **Marianne Lyons,** U.S. Navy Director, Acquisition Talent Management (DATM) **David Slade,** U.S. Air Force Director, Acquisition Career Management (DACM) **Otis Lincoln,** 4th Estate Director Acquisition Career Management (DACM)

Brigadier General Mike Sloane, USA (ret.)—is the Dean of the Defense Acquisition University's Defense Systems Management College (DSMC). He directly supports DAU's mission to provide a global learning environment to develop and assist senior acquisition, requirements, and contingency professionals who deliver and sustain effective and affordable warfighting capabilities.

Under Mr. Sloane's leadership, the DSMC faculty develop and manage dynamic curricula, teach, and facilitate learning, and continuously mentor course participants through the Executive Programs, Requirements Management, International Acquisition Management, and Leadership Learning Centers. He also leads Mission Assistance and Executive Consulting for senior leaders to solve complex challenges. Additionally he provides one-on-one Executive Coaching thru highly experienced faculty members.

Prior to joining DSMC in 2021, Mr. Sloane served in a Senior Executive Service position as a Deputy Director for the Counter Measures Acceleration Group (CAG), formerly Operation Warp Speed. He was responsible for providing strategic level Department of HHS/DOD nested pandemic planning capabilities for vaccines and therapeutics, a $30B/year mission. Additionally, he led the complex inter-agency DOD to HHS transition of the CAG mission to the Dept. of HHS w/no degradation in capabilities.

Mr. Sloane (Army Brigadier General, Retired) served over 29 years on active duty; 10 years of operational experience plus 19 years of defense acquisition experience in executive leadership positions. He served twice as a Program Executive Officer, first as PEO Simulation, Training, and Instrumentation (STRI) then as PEO Intelligence, Electronic Warfare and Sensors (PEO IEW&S). He also served 18 months as the Assistant PEO for Army Enterprise Information Systems. Prior to serving as the Asst PEO, he served in the Pentagon as the Chief of Staff for the Assistant Secretary of the Army, Acquisition, Logistics, and Technology.

Other key acquisition leadership positions include PM Soldier Sensors, Lasers, and Precision Targeting Devices; Product Manager Soldier Clothing, Equipment & Parachutes; Personnel Policy Integrator HQDA, Pentagon; APM System Level THAAD Test and Integration, and as APM Missile Development. He served as an Assignment Officer and in several operational assignments including Commander and Brigade Staff Officer, 10th Mountain Division, and Platoon Leader and Executive Officer, 24th Infantry Division. He has completed numerous combat and operational overseas tours. Mr. Sloane and Mrs. Sloane owned a Limited Liability Corporation in Greenville, SC which won the Small Business of the Year Award in 1997.

He holds several degrees to include a Master of Science, a Master of Business Administration, and a Bachelor of Business Administration. He is a Senior Service College and Army Command & General Staff College graduate, a Senior Acquisition Course graduate, and has completed the Executive Program Manager's Course at DAU. On active duty, he earned the Distinguished Service Medal, five (5) Legion of Merit awards, numerous other awards, and badges to include the Ranger, Airborne, and Air Assault badges.

Ronald R. Richardson, Jr.—currently serves as the Director of the Army Acquisition Support Center. In this role, he oversees the Army Acquisition Corps (AAC) and the Army Acquisition Workforce (AAW), and supports the Army's Program Executive Offices in the areas of human resources, resource management, program structure, acquisition information management, and program protection.

Mr. Richardson has over 30 years of medical, information, and weapon system acquisition experience as both a Department of Defense (DoD) civilian and a U.S. Army Officer. Before coming to ASC, he served as the Director of Acquisition and Operations for Program Executive Office Soldier. Prior to joining PEO Soldier, he was the Deputy Project Manager for the DoD Healthcare Management System Modernization (DHMSM®) Program, a $14B Major Automated Information System (MAIS) acquisition to replace the legacy Military Health System (MHS) Electronic Health Record (EHR) with an off-the-shelf (OTS) system now known as MHS GENESIS. Before that, he was the Product Lead for Increment 3 of the Integrated Electronic Health Record (iEHR) Program in the DoD/Department of Veterans Affairs Interagency Program Office (IPO). Prior to joining the DoD/VA IPO, he served as the Director of Acquisition Review and Analysis for the Office of the Assistant Secretary of the Army, Acquisition, Logistics and Technology (ASA(ALT)). Before joining ASA(ALT), Mr. Richardson served in a multitude of Military, Civilian, and Private Sector positions culminating in his selection for Senior Service College.

Mr. Richardson received his M.S. in Biomedical Engineering from Duke University, and his M.S. in National Resource Strategy from the Industrial College of the Armed Forces (ICAF). He is also a graduate of the U.S. Army Command and General Staff College.

He is the recipient of the Superior Civilian Service Medal (3), the Meritorious Civilian Service Medal (2), the Civilian Service Achievement Medal, the Army Staff Identification Badge, and the Order of Military Medical Merit (O2M3). Mr. Richardson also holds multiple professional memberships and certifications, including membership in both the Army and Defense Acquisition Corps, and Level III Defense Acquisition Workforce Improvement Act (DAWIA) Certification in Program Management, Science and Technology Management, and Systems Engineering.

Marianne Lyons—Since April 2019 Ms. Lyons has served as the Department of the Navy Director, Acquisition Talent Management (DATM). She is the Navy and Marine Corps' lead for the professional development and management of the DoN's over 70,000 civilian and military acquisition workforce. Ms. Lyons is the chief advisor to the Assistant Secretary of the Navy for Research, Development, and Acquisition, and guides all matters relating to initiatives and other strategic efforts that improve the acquisition workforce through education, training, and career management. She began her career with the Navy in 1989 as a naval architect and progressed to ship design management. In 2003, she transitioned to Program Management and later became an Action Officer at the Office of DASN Ships for the Auxiliary and Amphibious Ships portfolio. Prior to the DATM she was the Deputy Program Manager for the LPD 17 Amphibious Transport Dock Ship Program in PEO Ships. Ms. Lyons has a Civil Engineering degree from Virginia Tech and a Masters in Business from the Florida Institute of Technology. She is PM Advanced and ETM Practitioner DAWIA certified.

David Slade—is the Director of Acquisition Career Management, Assistant Secretary of the Air Force for Acquisition (SAF/AQH). Mr. Slade is responsible for the integrated management of the acquisition workforce across all functional areas. He provides acquisition human resources policy and strategic planning while managing the training and development of civilian and military acquisition personnel Air Force-wide. Additionally, Mr. Slade ensures Air Force compliance and implementation of the Defense Acquisition Workforce Improvement Act (DAWIA) through management of the Acquisition Professional Development Program (APDP) and the Defense Acquisition Workforce Development Fund (DAWDF). Mr. Slade is also designated as the Career Field Manager for both military and civilian Scientists, Engineers, and Acquisition Program Managers. His team also provides personnel management services for the SAF/AQ Headquarters Staff.

Mr. Slade received an aerospace engineering degree from the University of Colorado and was commissioned through the Reserve Officer Training Corps in 1983. Following pilot training, he served as a forward air controller, flying the O2-A and an F-15C and AT-38 instructor pilot. He served as a Commander at the Squadron and Group levels. As a command pilot with over 3,600 flying hours, he flew 32 missions over Iraq during Operation DESERT STORM and has participated in Operations NOBLE EAGLE, NORTHERN WATCH and SOUTHERN WATCH.

Prior to his current assignment, Mr. Slade served as Director of Assignments, Headquarters Air Force Personnel Center, Randolph Air Force Base, TX. He was responsible for the assignment of more than 65,000 officers below the grade of Colonel and 285,000 enlisted personnel below the grade of Chief Master Sergeant.

Mr. Slade retired from active duty, in the rank of Colonel, after 29 years in November 2012 and entered Civil Service in January 2013.

Otis R. Lincoln—entered federal service in 2009 as a Contract Specialist within the Office of the Chief Financial Officer (CFO) of the Defense Intelligence Agency (DIA). After serving as a Contract Specialist and a warranted Contracting Officer on several procurements supporting multiple Directorates across DIA, he continued to expand his aperture within the acquisition community moving into the project and program management realm. In multiple capacities, he was responsible for the successful planning and execution of various multi-million dollar programs that included increasing acquisition exposure to industry, training and career development of the agency's acquisition workforce as well playing an integral part of the hiring and placement of new acquisition members and set career paths in the finance and acquisition field. Mr. Lincoln has also utilized his Defense Acquisition Workforce Improvement Act (DAWIA) expertise in support of the Navy Systems Management Activity (NSMA) having served as their DAWIA Program Director overseeing and managing their workforce by expanding their training, certification, and career development. Following his tenure at NSMA, Mr. Lincoln assumed a senior leadership position as a Section Chief in the Contracting Office within CFO supporting the Mission Service's and Command Element's global procurement requirements. Currently, he serves as the Director, Acquisition Career Management for the 4th Estate (32 defense agencies/field activities) with oversight of statutory training, professional credentialing, continuous learning, and career development for more than 31,000+ acquisition workforce members.

PANEL 15. IMPLICATIONS OF THE NATIONAL DEFENSE STRATEGY FOR DEFENSE ACQUISITION SYSTEM

Thursday, May 12, 2022	
9:40 a.m. – 10:55 a.m.	**Chair: Cynthia Cook,** Director, Defense-Industrial Initiatives Group and Senior Fellow, International Security Program, Center for Strategic and International Studies **Panelists:** **David Arthur,** Weapon Systems Unit Analyst, Congressional Budget Office **Todd Harrison,** Senior Fellow, International Security Program and Director, Defense Budget Analysis and Aerospace Security Project, CSIS ***Defense Acquisition Trends 2022: A Preliminary Look*** Gregory Sanders, Center for Strategic and International Studies Alexander Holderness, Center for Strategic and International Studies

Cynthia Cook—is the director of the Defense-Industrial Initiatives Group and a senior fellow in the International Security Program at the Center for Strategic and International Studies. Her research interests include defense acquisition policy and organization, the defense industrial base, new technology development, and weapon systems production and sustainment. From 1997 to 2021, Dr. Cook worked at the RAND Corporation, where she oversaw, led, and worked on a wide range of studies for components across the U.S. Department of Defense, along with the Australian Department of Defense and the UK Ministry of Defense. Notable research leadership includes support for the government of Puerto Rico in the development of their congressionally mandated hurricane economic and disaster recovery plan and the publication of 20 supplementary analytic reports on recovery in every sector of the island's society. She led the implementation research for the 2010 report Sexual Orientation and U.S. Military Personnel Policy , which contributed to the elimination of Don't Ask Don't Tell. Her RAND management jobs include terms as the associate director of Project AIR FORCE and as the director of the Acquisition and Technology Policy Center in the National Security Research Division. Previously, Dr. Cook was a research specialist at the Massachusetts Institute of Technology, working on the Lean Aerospace Initiative. Before her graduate studies, Dr. Cook worked in New York as an investment banker, specializing in high-yield finance. She holds a PhD in sociology from Harvard University and a BS in management from the Wharton School of the University of Pennsylvania.

David Arthur—is a Senior Analyst in the National Security Division of the Congressional Budget Office. His work has focused on the budgetary aspects of long-term defense plans and the cost effectiveness of new weapon systems. During his 20 years at CBO, he has published reports on topics including combat aircraft, mobility systems, amphibious operations, missile defense, and the implications of budget constraints on defense planning. Prior to joining CBO, David was a research staff member at the Institute for Defense Analyses. He holds a PhD in physical chemistry from Stanford University, and bachelor degrees in chemical engineering and materials science and engineering from the University of California, Berkeley.

Todd Harrison—is the director of Defense Budget Analysis and director of the Aerospace Security Project at CSIS. As a senior fellow in the International Security Program, he leads the Center's efforts to provide in-depth, nonpartisan research and analysis of defense funding, space security, and air power issues. He has authored publications on trends in the defense budget, military space systems, threats to space systems, civil space exploration, defense acquisitions, military compensation and readiness, and military force structure, among other topics. He teaches classes on military space systems and the defense budget at the Johns Hopkins School of Advanced International Studies.

Defense Acquisition Trends 2022: A Preliminary Look

Greg Sanders—is a Fellow in the International Security Program and Deputy Director of the Defense-Industrial Initiatives Group at CSIS, where he manages a research team that analyzes data on U.S. government contract spending and other budget and acquisition issues. In support of these goals, he employs SQL Server, as well as the statistical programming language R. Sanders holds a master's degree in international studies from the University of Denver, and he holds a bachelor's degree in government and politics and a bachelor's degree in computer science from the University of Maryland. [gsanders@csis.org]

Alexander Holderness—is a research assistant with the CSIS Defense-Industrial Initiatives Group. He writes and researches on issues relating to national security, government acquisition, concept development, and industrial capacity. Prior to joining CSIS, Holderness worked as an intern for the U.S. Army Futures Command, Joint Army Concepts Division. He holds a BA in government and history from the College of William & Mary.

Abstract

This report is the latest in an annual series examining trends in what the U.S. Department of Defense (DoD) is buying, how the DoD is buying it, and from whom the DoD is buying. Fiscal Year (FY) 2021 proved to be the end of a five-year bounce back in defense contract spending, with contract obligations dropping to $380.1 billion, a 10% decline from FY2020 but still 28% higher than the FY2015 trough. This year's study focuses on the first year to partially fall under the new administration and examines how present trends align with the newly released National Defense Strategy fact sheet (DoD, 2022). The new administration has maintained a concern with speeding force development and technological adaption that justifies a continued focus on research and development in both contracting and other transaction authority (OTA) agreements. Additionally, this report includes analysis of the topline DoD contracting trends with particular attention to the report on the State of Competition within the Defense Industrial Base.

Introduction

This paper explores trends in defense acquisition in the transition year of Fiscal Year (FY) 2021, the end of a half decade of growth in defense contract obligations and the first to be partially administered by the Biden administration. The paper will look at two overarching questions: How have purchasing priorities shifted? How has the approach to the industrial base shifted? In answering these questions, the paper looks to the 2022 National Defense Strategy (DoD), which was released subsequent to this data and at the time of this writing is only publicly available in two-page factsheet form.

In addition, this analysis is informed by the recent report on the State of Competition within the Defense Industrial Base, which outlines five recommendations for increasing competition within the defense industrial base: strengthening merger oversight, addressing intellectual property limitations, increasing new entrants, increasing opportunities for small business, and implementing sector-specific supply chain resiliency plans (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2022, pp. 1–2).

This report uses the methodology employed in a range of CSIS reports on federal contracting. For over a decade, the Defense-Industrial Initiatives Group (DIIG) has issued a series of analytical reports on federal contract spending for national security by the government. These reports are built on Federal Procurement Data System (FPDS) data, which is downloaded in bulk from USAspending.gov, and, for other transaction authority data, from SAM.gov. DIIG now maintains its own database of federal spending, which includes data from 1990 to 2021. This database is a composite of FPDS and DD350 data. All dollar figures are in constant FY2020 dollars, using Office of Management and Budget (OMB) deflators. For additional information about the CSIS contracting data analysis methodology, see https://github.com/CSISdefense/Lookup-Tables.

Budgetary Context

The total obligation authority for the Department of Defense (DoD) declined from $725.8 billion in FY2022 to $697.5 billion in FY2021, a 4% reduction. These drops marked the end of a five-year bounce back in both defense budgets and contract obligations that started in FY2016 with the budget cap induced trough in FY2015. As shown in Figure 1, the decline was larger for contract spending with FY2021 with only $380.1 billion in obligations, representing a 10% drop from the $421.3 billion obligated in the prior year.

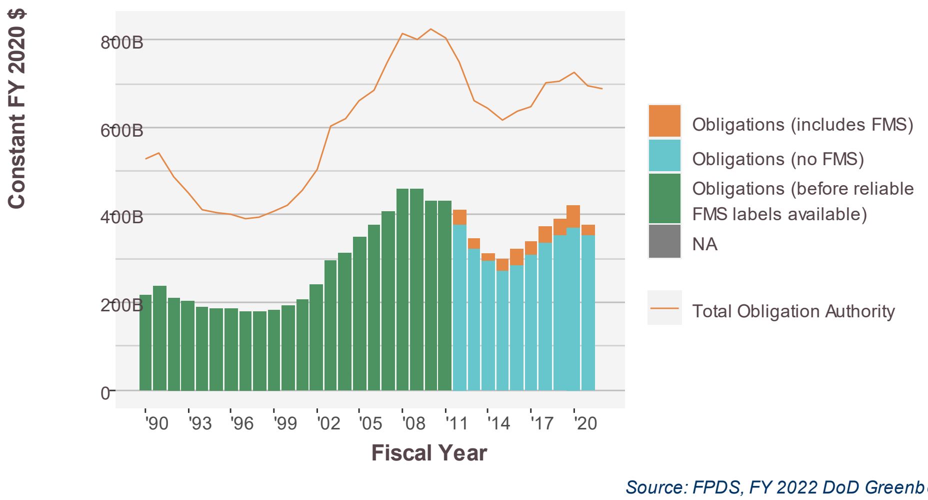


Figure 1. Defense Contract Obligations Versus Defense Total Obligation Authority, FY1990–FY2021

Defense contract obligations during this bounce back had steadily grown faster than defense budgets, peaking in FY2020 where the ratio of defense contract obligations to total obligation authority was the highest through the past three decades. This pattern flipped in FY2021, with contract spending falling faster than the budgeted obligation authority. One factor in this rise and fall was contracting for foreign military spending (FMS), which uses the U.S. defense acquisition system but relies on funding by international allies and partners or U.S. security assistance rather than the traditional U.S. defense budget.[1] This measure does overestimate total FMS because the contracts can also include a portion for use by the U.S. government. Nonetheless, it's remarkable that the portion of defense contract obligations with some FMS funding declined from $50.3 billion in FY2020 to $24.4 billion in FY2021, a reduction by more than half. One factor in the decline is that FY2020 included major multi-year contracts for systems with international components, most notably the F-35 Joint Strike Fighter. That said, the incoming Biden administration also did place a hold on some transfers to Saudi Arabia and the United Arab Emirates in an attempt to place pressure on those nations to bring an end to the war in Yemen.

[1] The Federal Procurement Data System provides enough information to identify the vast majority of contracts that include FMS spending, starting in FY2012. This is made possible by a field that directly identifies foreign funding and by reliance on treasury account code reporting that allows for the identification of foreign funding accounts.

The crisis caused by the unprovoked invasion of Ukraine by Russia has helped inspire increases in both the U.S. defense budget and agreements for arms transfers in FY2022, so the FY2021 decline likely does not herald the start of a new drawdown. Moreover, the challenging presidential transition that overlapped with the start of FY2021 also slowed the ability of the incoming administration to put their priorities into action. Thus this FY2021 should be interpreted as an important transition year and reset, but not necessarily indicative of trends to come even before accounting for shifts in response to the war in Ukraine.

Another notable shift in topline spending occurred in FY2021 as the five-year nearly exponential rise in other transaction authority (OTA) spending came to an end. OTA spending has an important role in adaptable acquisition framework reform efforts and was mentioned in the defense competition report alongside commercial solution opening approaches as a means of bringing in new entrants to defense acquisition (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2022, p. 2). As shown in Figure 2 in the blue line, OTA spending dropped to $15.1 billion in FY2021, a 7% decline from the $16.22 billion spent in FY2020. That said, the apparent doubling from FY2019 to FY2020 was already misleading, as in both of the two most recent fiscal years, the U.S. government response to Covid-19 drove at least $10 billion in OTA spending. This is in keeping with pandemics being identified as one of the "transboundary threats" mentioned in the 2022 National Defense Strategy. OTA spending will continue to merit close examination to see whether spending levels have reached a plateau or whether the remarkable rise in base and all options value shown in the orange line presages further notable increases. Because OTA data is only reliably available starting in FY2015, it is not included in the graphs and figures for the remainder of the report except where explicitly noted.

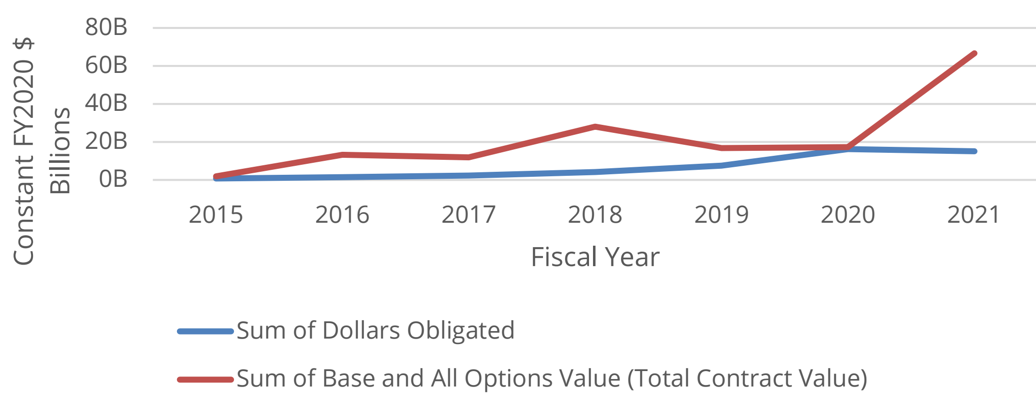


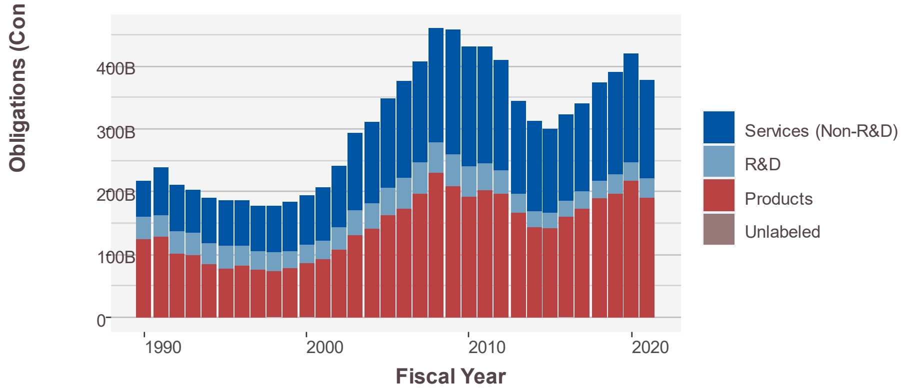
Figure 2. Defense OTA Spending, FY2015–FY2021

How Have Purchasing Priorities Shifted?

FY2021 saw a return to a more traditional share of obligations going to products spending, which fell from $218 to $191 billion. As will be covered in subsequent sections, much of this decline and FY2020's heights can be attributed in part to major weapon system contracts that cover multiple years and have uneven spending patterns. Even with the whipsawing, products still accounted for a little over half of DoD contract obligations (50.2%). Service spending also declined, accounting for $158.2 billion in FY2021, 9% below the prior year's $173 billion.

In keeping with the National Defense Strategy fact sheet's emphasis goal to "accelerate force development," the one category that increased in FY2021 was R&D, which rose to $30.9 billion in FY2021, a 2% increase over the $30.2 billion in obligations in FY2020. While R&D

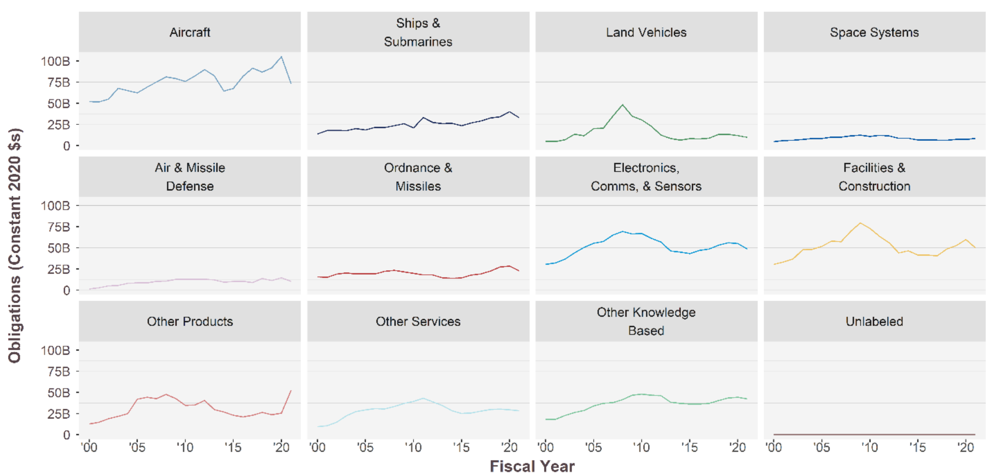
contract spending has still grown slightly slower than defense spending overall since the FY2015 trough, 26 and 28% respectively, as is covered in Figure 3, this contract spending is supplemented by significant growth in OTA R&D expenditures over the same period.



Source: FPDS; CSIS analysis

Figure 3. Defense Contract Obligations by Product, Service, and R&D, FY1990–FY2021

The DoD has been making hefty investments across the defense enterprise; however, in FY2021 the overall 10% decline reflected funding decreases across almost all platform portfolios. As seen in Figure 4, over the last several years spending on aircraft and ships and submarines has hit record highs. In 2021, those two leading categories have declined by 30% and 18% respectively, though this reflects that their procurement processes often involve lumpy orders with obligations for contracts lasting multiple years peaking in FY2020 and covering the spending valley FY2021. Given the FY2023 budget, the services modernization priorities, and the war in Ukraine, this decline is unlikely to be sustained. Ordnance and missiles is an important category to watch—even after declining 20% in FY2021 it is still the weapons system category that has grown the most since the trough in FY2015, increasing by 59%. Nonetheless, replacing defense articles given to Ukraine and the development and deployment of hypersonic missiles and other advanced ordnances will likely see the category growing again.



Source: FPDS; CSIS analysis.

Figure 4. Defense Contract Obligations by Platform Portfolio, FY1990–FY2021

Only two portfolios experienced notable growth: other products which more than doubled, increasing 106% largely to cover expenditures relating to the government's Covid-19 response. More pertinent to traditional defense acquisition, space system obligations increased by 19% in FY2021, growing from $7.2 to $8.6 billion. In previous years, CSIS had noted that this category was growing slower than the overall defense spending increases despite being one of the categories emphasized in the prior national defense strategy, likely partially because the space spending is thought to have a significant classified component.

Spending across the services remained largely flat, as shown in Figure 5. The 8% FY2021 increase in Army spending was driven largely by the Covid-19 response effort, which the Army contracted. Additionally, Army spending air and missile defense as well as ordnance and missiles decreased in FY2021, which does not align with the predicted shifts in spending as the service works to operate in a higher-level threat environment. Air Force spending has remained largely flat, with a decline that rounds to a 0% change, despite sizeable investments in the B-21 program and Ground Based Strategic Deterrent. Air Force spending on classified programs (including NGAD in FY2021) may have resulted in an overall increase in spending. Despite flat topline spending, the Air Force does seem to be making sizable investments in order to return to an era of great power competition.

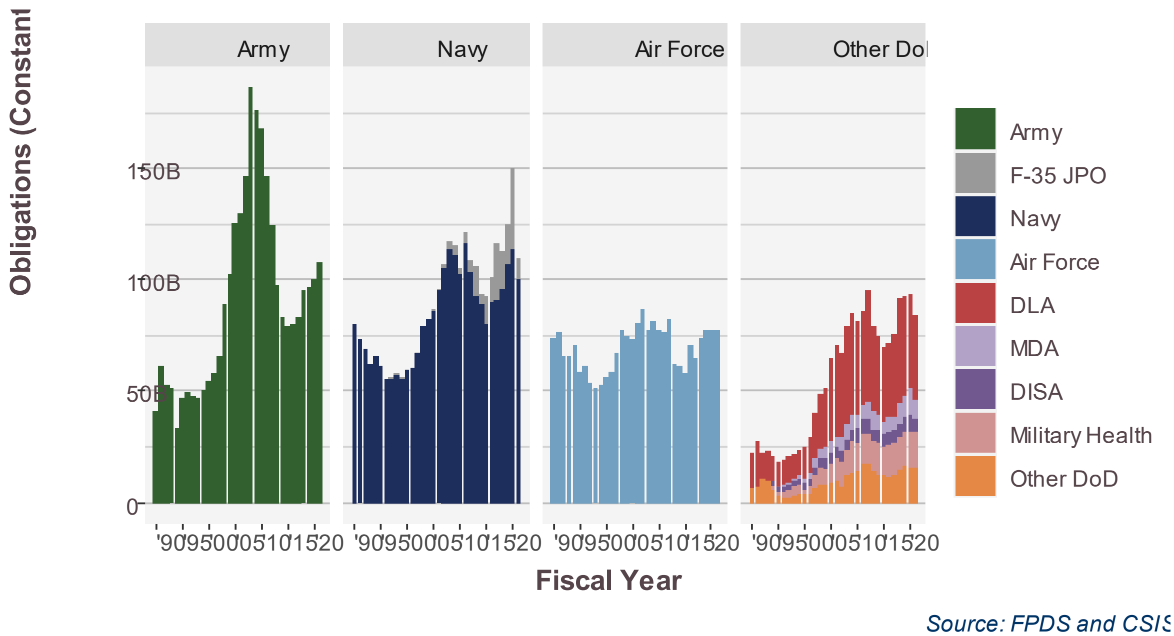


Figure 5. Defense Contract Obligations by DoD Component, FY1990–FY2021

Turning to other defense agencies, the Defense Logistics Agency is the largest contract spender and declined to $38.5 billion in FY2021, an 8% drop. The military health agencies, including Tricare and the Defense Health Program, were an exception to the larger decline, growing to $15.7 billion in FY2021, a 2% increase. In contrast, in FY2021 the Missile Defense Agency fell to $8.7 billion and the Defense Information System Agency fell to $5.9 billion, declines of 30% and 23% respectively, both faster than the overall decline.

Navy spending declined in FY2021 for a variety of reasons, though that decline in spending does not necessarily mean that the Navy is seeing cuts to its capabilities, operational tempo, or readiness. First, in FY2020 the F-35 JPO placed an order for a low rate initial production lot of F-35s, an order that happens every few years. That order, which covers several years of procurement, did not repeat in FY2021, resulting in a one-year spending drop from $35.1 to $9.7 billion. The decrease in Navy spending that is not related to the F-35 JPO is more complex. In FY2021 Navy spending on ships and submarines decreased. This is in part because the Navy authorized long lead time procurement on several submarines and ships in FY2020 that covered multiple years of procurement, meaning the cost that covers multiyear procurement will show up unevenly in FY2020. Across all other platform portfolios spending remained largely flat, making it probable that "lumpy" contracting is largely to blame for the Navy's decline in spending, and not a strategic decision to underinvest in the service despite renewed focus on the Indo-Pacific region.

Reflecting the report's emphasis on rapidly deploying technology to military operators, spending on software specifically is worth a closer look. Figure 6 shows prime contract spending for software, which will predominantly fall in the electronics, comms, and sensors platform portfolio. This represents only a portion of total DoD software spending, as much of the effort goes towards software embedded in larger weapon systems that is not broken out in separate contracts. Nonetheless, software acquisition has been a perennial challenge for the DoD and it is worth taking a closer look at changes in the extent of direct purchase and forms of acquisition.

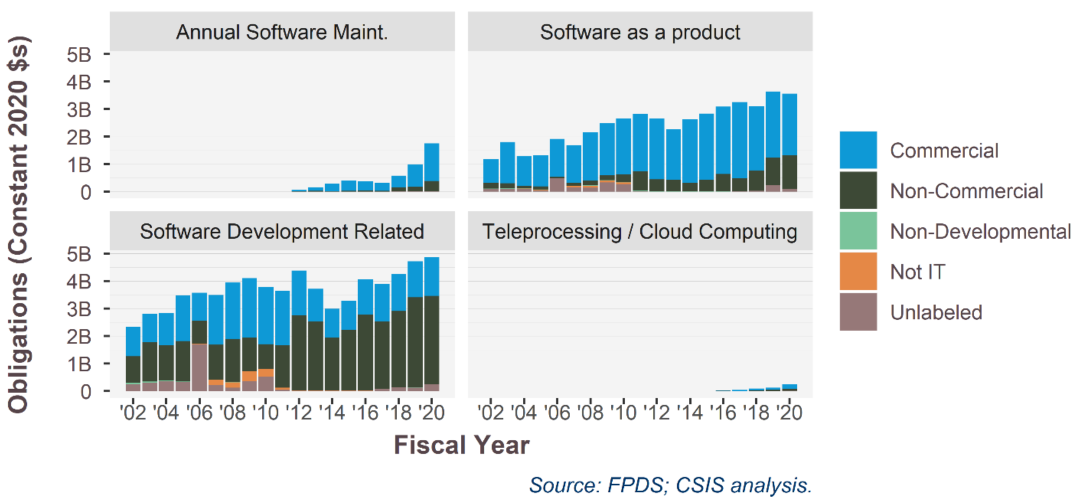


Figure 5. Defense Contract Obligations for Software by Commercial Status, FY2002–FY2020

Unsurprising, in FY2020 spending on prime contracts for software continued to increase across the DoD, growing to $10.4 billion, a 11% increase. Two increases in spending are critical to understanding the future of IT acquisition: annual software maintenance and cloud computing. Historically, the DoD has bought much of its commercial software, shown in sky blue, as a product that included a permanent license but would need to be replaced as it went obsolete. In FY2020 the DoD spent $3.6 billion on software as a product, a 1% decline from FY2019. Annual software maintenance is equivalent to the commercial software as a service model. Contracting software as a service allows agencies to upgrade their capabilities, patch vulnerabilities, and better manage costs. The DoD spent $1.75 billion on annual software maintenance in FY2020, less than the software as a product or paying for software development directly, but an 81% increase over the prior years. FY2020 spending on teleprocessing and cloud computing still has a low baseline of only $0.26 billion dollars in FY2020, but this amount represents a doubling over the previous year. While this is not evidence of a new acquisition approach, it demonstrates that the department is serious about developing cloud solutions for both warfighters at the tactical edge and support agencies managing vast data and physical enterprises. For more on these information technology issues, see *Leveraging Networks in Future Operations* by Gregory Sanders and Rhys McCormick (2022).

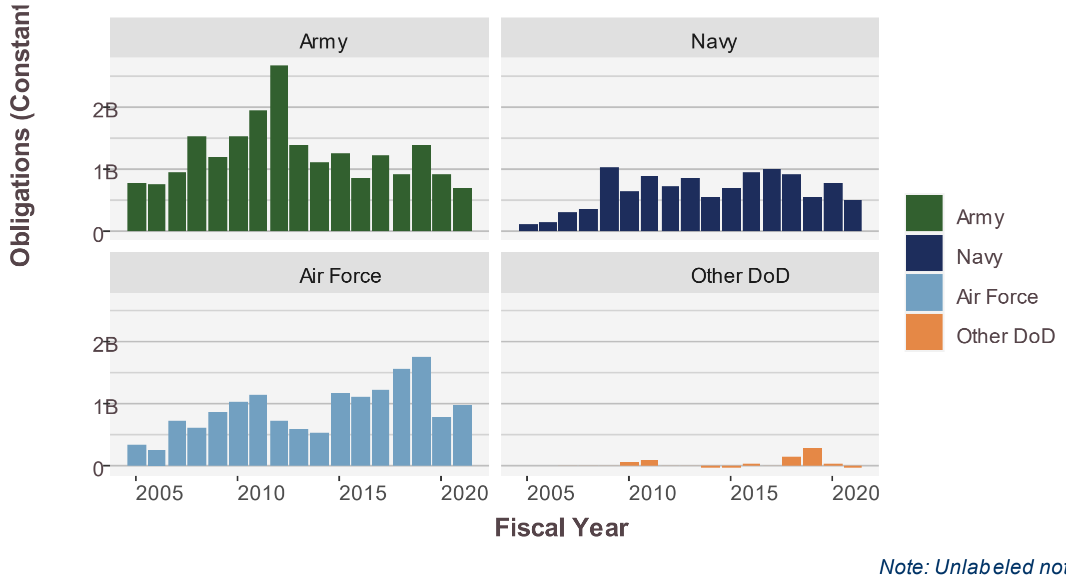


Figure 6. Defense Contract Obligations for Remotely Crewed Systems, FY2005–FY2021

Remotely crewed systems have long offered operation advantages to the department and to operators. These advantages, which include decreased operational costs and lower risks to personnel, have led the department to spend billions on the systems. Remotely crewed systems are difficult to track in FPDS, but Figure 7 is a first attempt to bring more transparency to this category. Unlike other systems that the DoD buys, uncrewed systems often include a higher level of commercial components. Sustained DoD spending over the last decade is partially indicative of greater acceptance of commercial solutions for defense applications that feature significant amounts of emerging technologies. That said, the comparatively flat spending in recent years, with only $2.5 billion in FY2021, indicates that to the extent new spending is happening, it is often either taking place in the classified space or is poorly captured by existing product or service classifications that focus on aerial systems as a product.[2] For a broader discussion on remotely crewed systems see *Reaching Farther, Risking Less,* a CSIS report by Rose Butchart and Gregory Sanders (2021).

How Has the Approach to the Industrial Base Shifted?

Increasing the speed of modernization and technological adoption is the most clear-cut objective for the acquisition system in the strategy. This is a continuation of the objectives of the larger adaptable acquisition framework. One way to measure speed is to look at the speed of individual projects, as Morgan Dwyer (2020) has done in an examination of cycle time and the centralization or decentralization of the acquisition system. Her findings, along with those of David Tate (2016), do raise questions about how achievable attempts at speed will be and the way they could be undermined in the absence of effective oversight or by setting ambitions such as ambitious software goals that may be incompatible with moving fast. Much of this literature focuses on major defense acquisition programs (MDAPs), both because these are the largest

[2] There is a single product or service category for Unmanned Aircraft (1550), but no codes for R&D or services related to remotely crewed systems or for the purchase of ground or maritime remotely crewed systems.

and typically most ambitious weapon systems but also because there is a wealth of reporting on the cost and schedule of these efforts. The middle tier of acquisition, which emphasizes the ability to deliver results in less than five years, does not have such rigorous reporting requirements, and is especially difficult to track using publicly available data. This is in part an intentional effort to avoid replicating some of the negative dynamics of the MDAP system, but it also makes it challenging to evaluate how the acquisition system is performing with regard to the strategy's objectives.

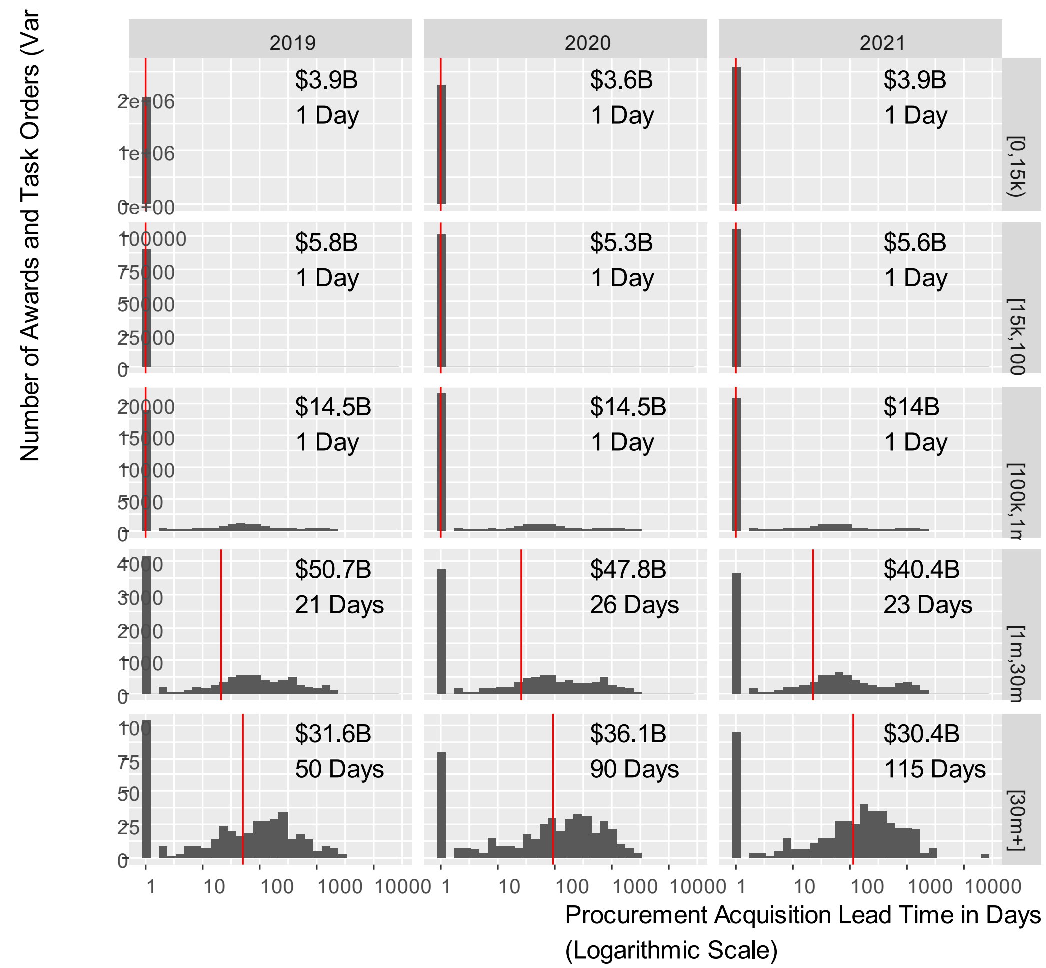


Figure 7. PALT by Start Year for Single-Award and Initial Contract Ceiling, FY2019–FY2021

Happily, reformers have provided another way to measure the speed of the acquisition system by mandating the development and reporting of a consistent definition of procurement award lead time (PALT; Berteau, 2018). PALT captures part of the lead time for the contracting system, the amount of time between the initial solicitation date and the award date for contract. PALT is not all encompassing; much time can pass between the desire for a good or service and solicitation, and of course delays can occur during the contracting process. Nonetheless, by establishing a consistent definition and then mandating the reporting of solicitation dates for DoD contracts above the simplified transaction threshold, reformers created a valuable tool for evaluating the agility of acquisition system and the magnitude of hurdles to getting projects under way (Assad, 2018).[3] One vivid example of why PALT matters is that the space launch industry will sometimes have unused capacity as launch time approaches, offering a chance at a dramatically cheaper way to put a ready satellite into space, but only if the contracting can be complete in advance of the launch date.

To evaluate efforts to speed the defense acquisition system, Figures 8 and 9 look at the distribution of PALT by contract start year, with each year being its own column, and by initial contract ceiling, with each bucket of range of contract ceilings reported in a separate row. Each row of these graphs uses a different y-axis scale, because there are orders of magnitude more smaller contracts than larger contracts. That said, as the dollar figures in each cell show, even though contracts with a ceiling above $1 million (shown in the bottom two rows of each graph) account for only thousands of the millions of awards and task orders given in these years, they also account for the majority of the obligations.

These graphs evaluate PALT by looking at the median number of days from solicitation to signed contract in each cost ceiling category for each start year.[4] Figure 8 reports on task orders for single-award indefinite delivery contracts (IDCs) as well as IDCs of unknown type. These vehicles are examined separately because these vehicles often pre-specify almost all aspects of a contract, allowing for straightforward award of awarding within a single day.[5] Single-award IDC task orders can include very complex tasks subject to significant negotiation, witness the wider range of PALTs in the lower rows of Figure 8; however the starting conditions of an existing contract with a single vendor makes them different enough to merit separate consideration from other contract types.

For single-award IDCs, the typical contracts with a ceiling under $1 million is executed in a day and the average PALT is less than 10 days. Higher ceiling single-award IDCs are more concerning, as both average and median PALTs are higher in FY2021 than in FY2019 and on the average PALT for task orders with ceilings of $30 million or higher rising to over a year and the median task order taking 115 days, an increase of 28% from FY2020 contracts.

For other contract types including definitive award and a range of multiple-award vehicles, PALT is getting worse for large contracts, but there are signs of improvements for some contracts with ceilings below $1 million. The signs of improvement for smaller contracts is corroborated by the competition report's finding that the DoD significantly improved on meeting the required 90-day notification to small businesses of decision to award as well as the Small Business Administration's recommended contract award times of 180 days from the close of the initial SBIR/STTR solicitation" (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2022, p. 15). The speed of awarding contracts with ceilings above $1 million had shown improvement in FY2020, only to backslide to longer median and mean value in FY2021, with contracts between $1 million and $30 million having a median PALT of 81 days and those above $30 million having a PALT of 241. The changes for contacts with ceilings

---

[3] Reporting was required as of June 19, 2018. This means that complete reporting is not available for FY2018 and thus Figures 8 and 9 start with contracts issued in FY2019 for greater comparability.

[4] The authors chose to use the median in part because of concerns over outlier solicitation dates. Over 200 contracts had solicitation dates that were at least 10 years old, and those included ranged back to October 1, 1957. Those prior to the establishment of DoD in 1947 were treated as input errors and removed from the data set.

[5] This may happen when "the action is the award of an order using existing pre-priced line items under an indefinite-delivery contract where no proposal is required (i.e. there are no elements to delivery or performance to negotiate)" (Assad, 2018, p. 1).

between $15,000 and $1,000,000 showed more consistent progress, with lower median and mean times, all now below 90 days.

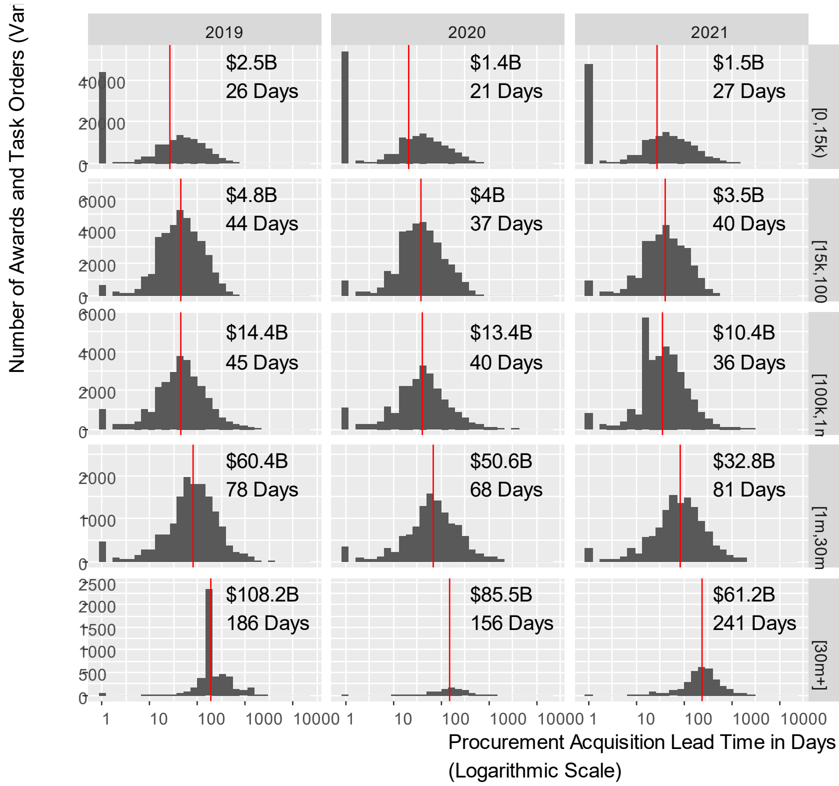


Figure 9. PALT for Other Awards and Task Orders by Start Year and Initial Contract Ceiling, FY2019–FY2021

While not covered directly in the strategy, competition has been an area of focus for the Biden administration and FY2020 had been a low point in competition shares for this century, driven by the procurement of major weapon systems such as the F-35 (Sanders et al., 2022).

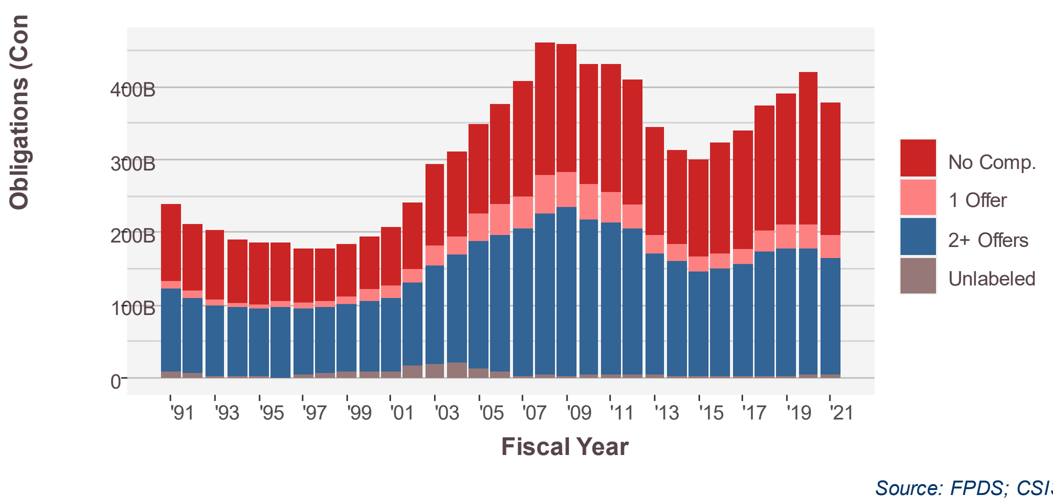


Figure 10. Defense Contract Obligations by Extent of Competition, FY1991–FY2021

In FY2021, competition for contracts across the department increased, but only 42% of obligations had been competed with two or more offers, an improvement but still the second lowest rate for the period shown in Figure 11. During that period, $182.8 billion went to contracts awarded without competition, a 13% decline. Obligations also declined for all forms of contracts awarded with competition, although the smallest decline was for those that had received only a single offer, which fell to $32.6 billion, a drop of 4%. The small move toward more competition is in line with the desired direction of the department's political leadership, but the report on the state of competition noted a range of concern with consolidation, intellectual property, and data rights, and in particular with a declining number of prime contractors in major weapon categories (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2022, p. 8). As shown in Figure 11, the rate of competition varies greatly between platform portfolios. In some cases the shifts can be clearly attributed to the movement of major weapons systems through their life cycle, with the joint strike fighter being a sufficiently large program to reshape the aircraft sector. Other cases, such as ordnance and missiles, are shaped by wider budgetary and consolidation decisions. The competition report calls out the missiles and munitions sector as one where competition is a concern:

> The growing pressure on defense budgets to reduce costs and spending has negative effects on munitions programs—including service cuts and congressional program reductions. While the budgets for munitions have not returned to their 2015 low, the services tend to flatten M&M procurements or cyclically push procurements into the out year. As commodity costs grow, these factors drive suppliers to exit the market rather than join it, such as automation solutions companies pivoting away from lower-margin defense programs. (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2022, p. 19)

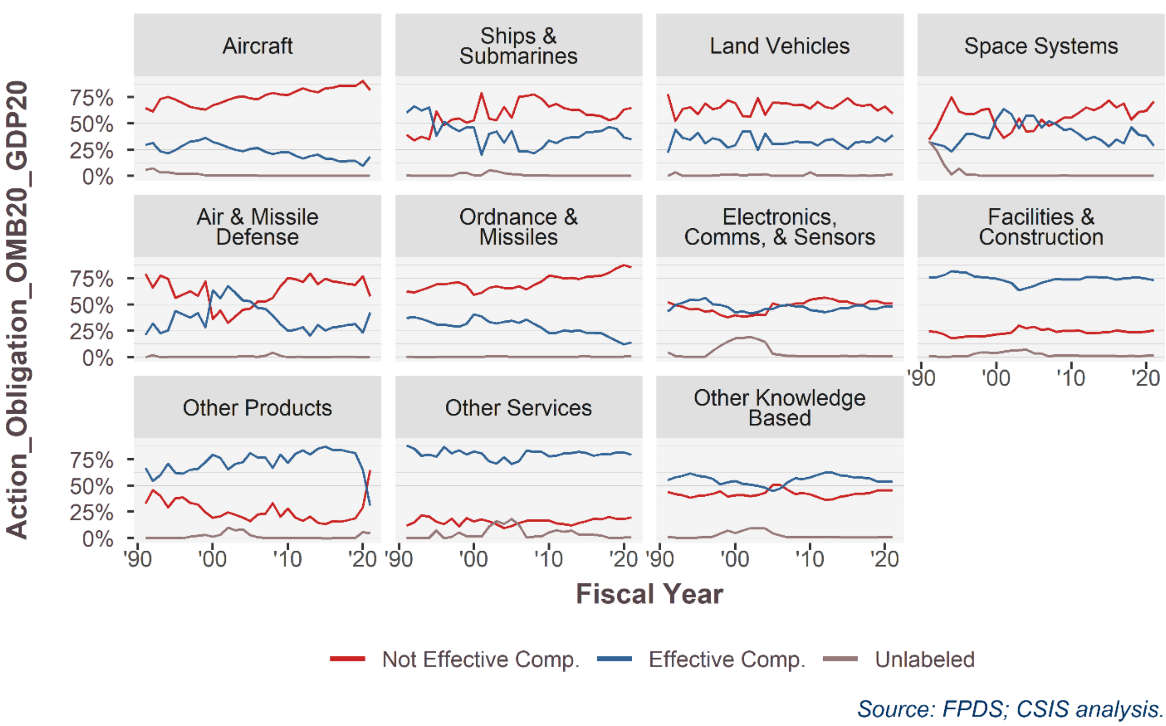


Figure 8. Competitive Market Share by Platform Portfolio, FY1991–FY2021

A lack of competition certainly hurts the DoD's ability to bring in innovative technologies while continuing to manage down procurement costs. As the DoD starts to have a full roster of political appointees confirmed, the department may be better able to mandate and drive effective competition, though those results will not start to be seen in spending and competition data until FY2022 or even possibly the FY2023 data.

Both the competition report and critics of the report such as David Berteau (2022) highlighted the challenges for new entrants to the acquisition system. The data underlines this concern—past CSIS research has found that the number of DoD new entrants had peaked in FY2005 only to steadily decline through FY2013 before holding relatively steady.[6] More recently, "the total number of prime vendors serving the DoD fell to below 41.6 thousand in FY2020, a 10% decrease" (Sanders et al., 2022, p. 7). Contracting policy may be contributing this decline, "DOD officials and small business executives told [the Government Accountability Office] the category management initiative reduces opportunities for small businesses, which may have difficulty participating in large government-wide contracts such as the initiative's Best in Class contracts" (Shear, 2021, p. 11). Figure 11 shows why this might be the case. Multiple-award IDCs and other indefinite delivery vehicles (IDV) constitute a growing share of obligations going to small and medium vendors. In FY2015, multiple-award IDCs and other IDVs constituted 33% and 21% of small and medium vendor obligations respectively; by FY2021 that portion rose to 39% for small vendors and 26% for medium vendors. The change is not so large as to suggest that it has played a decisive role in influencing the number of new entrants, although it does suggest there is merit in the goal listed in the competition report to "create more opportunities annually for small businesses to onboard onto contract vehicles and compete for contract awards" (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2022, p. 16). That said, greater success in allowing small vendors to the on ramps to these vehicles would not address concerns raised by Berteau (2022) that "a small business can win contracts and become too large for set-aside programs, but DoD offers too few full-and-open competitions for those just-graduated companies."

[6] This peak can be partially explained by a lowering of the reporting threshold from $25,000 to $2,500 at roughly the same period. However, while the reporting of new entrants that had previously fallen under the threshold does explain why a peak occurred in FY2004 and FY2005, but does note explain why the number of new entrants steadily dropped in the years thereafter.

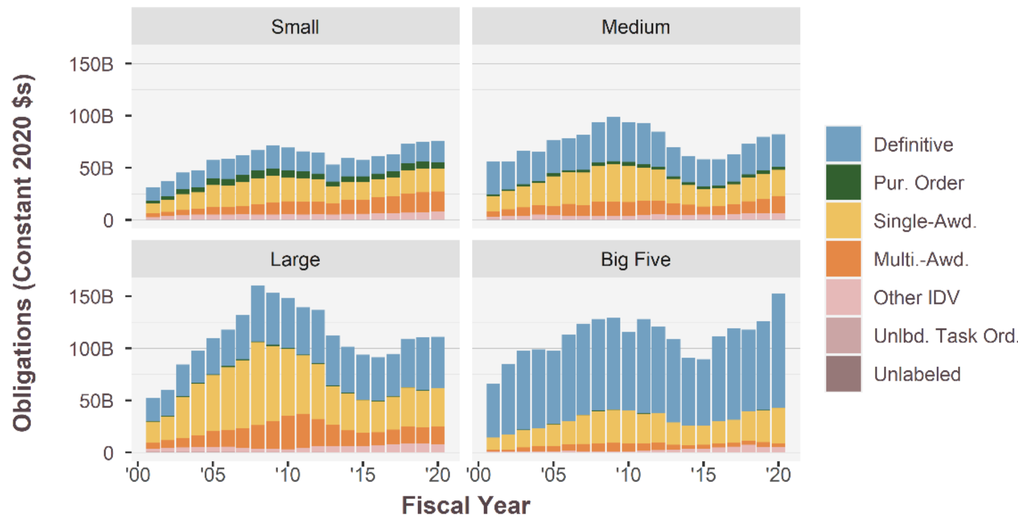


Figure 12. Defense Contract Obligations by Vendor Size and Vehicle, FY2000–FY2021

One approach the DoD has taken to increasing competition over the past few decades has been greater reliance on commercial contracting, which has the option to call on less restrictive acquisition requirements due to reliance on greater competition in commercial markets to ensure performance. The competition report cited a notable increase in use of these procedures:

> According to DoD contract award data from the Federal Procurement Data System, the early 2000s saw commercial items make up 30−50% of all procurements. Since 2011, commercial items have consistently accounted for over 88% of new awards (and as high as 98% of new awards) across DoD. (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2022, p. 12)

As shown in Figure 12, the share of dollars is notably lower than the share of awards, with only 26% of obligations, $99 billion in FY2021, classified as commercial according to CSIS analysis.[7] However, there was a striking increase in the use of commercial contracting in FY2021, with a 37% increase over the $72.4 billion spent in FY2020. Throughout the FY2015–FY2020 period, obligations for commercial products and services grew only 32%, less than the 42% overall rise in defense contract obligations. This FY2021 spike has can be entirely explained by the Army's purchase of other products in response to the Covid-19 epidemic.

[7] The study team included any transaction that used any form of commercial acquisition procedures or that was classified as a commercial information technology. Both commercial products and services, to the extent they are described as such in FPDS, are captured in this analysis.

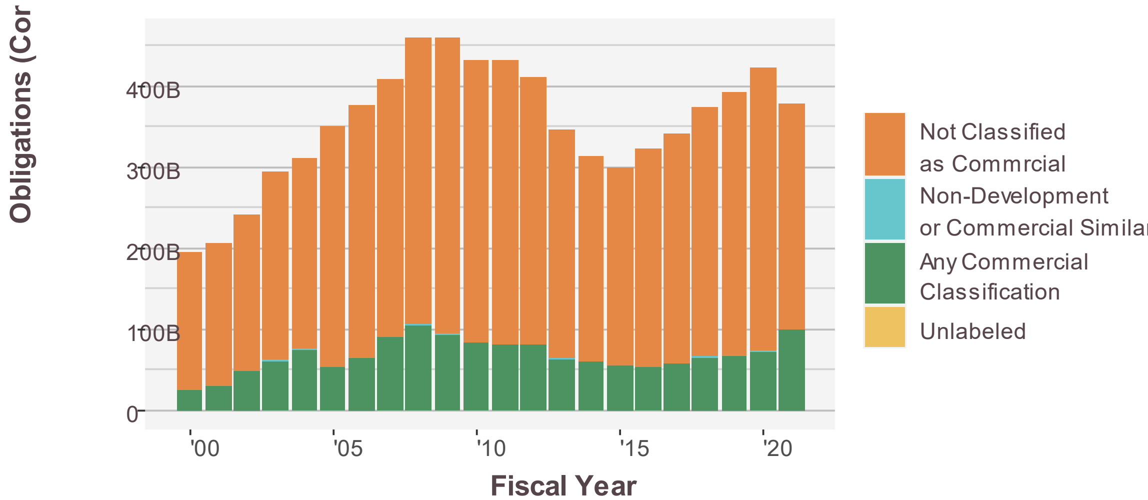
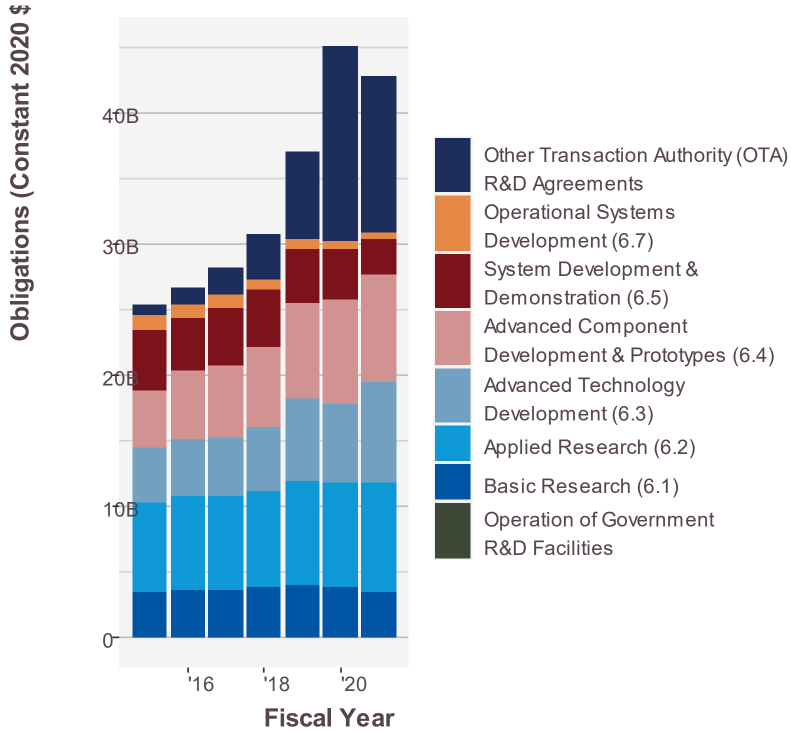


Figure 13. Defense Contract Obligations for Commercial Items and Services

Finally, one major approach to bringing in commercial technology not covered by Figure 11 is the use of OTA. As covered in Figure 2, OTA expenditures declined by 7% in FY2020, a slightly slower rate than the 10% decline in traditional defense contracting. However, in R&D specifically, contract spending grew by 2% while OTA R&D expenditure dropped by 19%, from $14.8 billion to $11.9 billion. Even so, as shown in Figure 14, OTAs still represent 28% of combined contract and OTA R&D expenditures, the second highest share of the study period. In both FY2020 and FY2021, Covid-19 was a significant part of OTA expenditures, a topic which will merit further study as unlike contract expenditures tracked in FPDS, the relevance of OTA transaction is not officially reported and thus researchers must manually classify contracts and transactions.

Focus on traditional R&D contracting, both early and late phase R&D faced cuts with basic (6.1) research declining by 9% and system development (6.5) and demonstration and operational system development (6.7), falling by 28% and 40% respectively. The largest growth was in advanced component development (6.3), which grew by 27%. Given the need for rapid modernization, the DoD has clearly made trade-offs as it tries to strike a balance between basic research that will yield revolutionary technologies in decades, and more incremental gains in capability that the Joint Force and DoD need to be able to compete against more assertive global pacing threats. The DoD has proven increasingly capable of getting to the prototyping stage (6.4) with both OTA and traditional contracts, but translating those prototypes into procurement programs remains a challenge.



Source: FPDS and CSIS analysis.

Figure 9. Defense R&D Contract and OTA Obligations, FY2015–FY2021

Conclusions

The bounce back after the drawdown and budget caps has reached its end

Spending has declined by 10% in FY2021 with a broad based decline in contract obligations, falling at a faster rate than the 4% fall in budgetary total obligation authority. FY2022 may yet see an increase, though the role of inflation means that even still growing budgets may not be enough to result in a net increase in obligations to industry.

Responses to Covid-19 shaped spending priorities using tools aimed at commercial technology

In line with the strategy's emphasis on responding to cross-boundary threats and integrating all tools of national power, the defense acquisition system supported the larger response to the present pandemic including notable jumps in Army spending for commercial products and FY2020's dramatic increase in OTA spending. This example may merit closer to study to learn what successes and failures might be applied to future national emergencies.

R&D and space systems stood out as protected priorities even as overall obligations declined

R&D contract spending grew by 2%, although that growth was offset by a 19% decline in R&D OTA spending. Spending for space systems was the one portfolio focused on weapon systems that grew, increasing by 19% to $8.6 billion.

Acquisition agility, as measured by PALT, is not improving for large contracts.

Three complete years of data on PALT is now available and unfortunately the story it tells does not show signs of improvement in FY2021 compared to the prior two years. Notably, the median length for a single-award IDC contract task order over $30 million is 115 days, with the average lead time being over a year (410.4 days). Complications to administration under

Covid-19 conditions may have contributed to these delays, but the absence of progress potentially threatens the goals laid out in the NDS strategy. Administrative improvements were enabled by large data set analysis such as that researched by David Gill and Timothy Hawkins (2021). Another key aspect of the problem is the efforts for Planning, Programming, Budgeting, and Executing reform, as even the most efficient acquisition vehicles cannot result in a signed contract unless the money is authorized to pay for it.

References

Assad, S. (2018). Reporting "solicitation date" in the Federal Procurement Data System [Memorandum]. DoD. https://www.acq.osd.mil/dpap/policy/policyvault/USA001458-18-DPAP.pdf

Berteau, D. (2018, March 30). DoD can reduce time to contract. Federal Times. https://www.federaltimes.com/acquisition/2018/03/23/dod-can-reduce-time-to-contract/

Berteau, D. J. (2022, February 16). The problems with the defense-industrial base aren't about the contractors. Defense News. https://www.defensenews.com/opinion/commentary/2022/02/16/the-problems-with-the-defense-industrial-base-arent-about-the-contractors/

Butchart, R., Sanders, G., Cheney, J. N., & Araz, S. (2021). Reaching farther, risking less: Remotely crewed systems' ISR applications for great power competition. Center for Strategic and International Studies. https://www.csis.org/analysis/reaching-farther-risking-less

DoD. (2022). Fact sheet: 2022 National Defense Strategy. https://media.defense.gov/2022/Mar/28/2002964702/-1/-1/1/nds-fact-sheet.pdf

Dwyer, M., Tidwell, B., & Blivas, A. (2020). Cycle times and cycles of acquisition reform [CSIS briefs]. Center for Strategic and International Studies. https://www.csis.org/analysis/cycle-times-and-cycles-acquisition-reform

Gill, D., & Hawkins, T. (2021). *How long does it take to award a government contract? Understanding PALT time frames with big data analytics*. Acquisition Research Program. https://dair.nps.edu/handle/123456789/4403

Office of the Under Secretary of Defense for Acquisition and Sustainment. (2022). State of competition within the Defense Industrial Base. DoD. https://media.defense.gov/2022/Feb/15/2002939087/-1/-1/1/state-of-competition-within-the-defense-industrial-base.pdf

Sanders, G., Jang, W. J., & Holderness, A. (2022, March 29). Defense acquisition trends 2021. https://www.csis.org/analysis/defense-acquisition-trends-2021

Sanders, G., & McCormick, R. (2022). Leveraging networks in future operations: DISA's changing role in battle networks.

Shear, W. B. (2021). Small business contracting: Actions needed to implement and monitor DoD's small business strategy. GAO. https://www.gao.gov/assets/gao-22-104621.pdf

Tate, D. (2016). *Acquisition cycle time: Defining the problem.* Acquisition Research Symposium.

PANEL 16. UNDERSTANDING REQUIREMENTS: PEOPLE, PROCESSES, AND TECHNOLOGY

Thursday, May 12, 2022	
9:40 a.m. – 10:55 a.m.	**Chair: John J. Hammerer,** Chair, Integrated Air and Missile Defense, Naval Postgraduate School ***Understanding Post-Production Change and Its Implication for System Design: A Case Study in Close Air Support During Desert Storm*** Aditya Singh, George Washington University Zoe Szajnfarber, George Washington University ***Correlating a User Experience System to Product Success*** Jonathon Miller, Naval Information Warfare Center

John J. Hammerer—is the Chair, Integrated Air and Missile Defense at the U.S. Naval Postgraduate School. Previously he served as a Senior Research Scientist at the Center for Naval Analyses, the Navy's Federally Funded Research and Development Center where his principle areas of research included combat systems development and operational warfighting assessment.

Prior to joining the Center for Naval Analyses, he was a defense consultant to the AEGIS and New Construction Ships Branch in the Office of the Chief of Naval Operations and the AEGIS Ballistic Missile Defense Program Office. In this capacity he was a member of several study groups including the Congressionally mandated BMDS Training and Education Needs Assessment and the DDG-1000 Combat System Activation Readiness Assessment.

John's eight shipboard assignments, included duty as the commissioning Commanding Officer of USS Paul Hamilton and command of USS Lake Erie, conducting the first ballistic missile intercepts in space from a ship at sea.

Ashore, he served as the Commander of the Ballistic Missile Defense Organization's Joint National Test Facility, the Director of the Missile Defense Agency's Initial Defensive Operations Task Force and Program Manager of the Ground Based Midcourse Defense Fire Control and Communications System. He also led the Ballistic Missile Defense Organization's Joint Force and Test and Evaluation of Battle Management Command, Control, and Communications, and Intelligence Directorate. John was the Surface and Strike Warfare Officer in Combat System Engineering Branch of the AEGIS Program Office's Technical Division.

Prior to reporting to the Naval Postgraduate School, he was an adjunct professor of Joint Maritime Operations at the U.S. Naval War College.

He holds a Bachelor of Science degree in Marine Science from the University of South Carolina, a Master of Science degree in Physics from the U.S Naval Postgraduate School and a Master of Science degree in National Security and Strategic Studies from the U.S. Naval War College. He was certified as a Joint Specialty Officer and Acquisition Professional Level III in Program Management.

Understanding Post-Production Change and Its Implication for System Design: A Case Study in Close Air Support During Desert Storm

Aditya Singh—is a PhD student at The George Washington University, studying Engineering Management in the Department of Engineering Management and Systems Engineering. He received his BS in Systems Engineering and Economics from The George Washington University and has worked at numerous government contractors including Lockheed Martin and BAE Systems Inc. Currently, he is a fellow on The George Washington University's National Science Foundation Traineeship Program: Designing Trustworthy AI Systems. [asingh25@gwu.edu]

Zoe Szajnfarber—is a Professor and Chair of Engineering Management and Systems Engineering at The George Washington University. Her research seeks to understand the fundamental dynamics of innovation in technology-intensive governmental organization, as a basis for decision-making. She received her bachelor's degree in Engineering Science from the University of Toronto. Szajnfarber conducted her graduate work at the Massachusetts Institute of Technology, earning dual master's degrees in Aeronautics & Astronautics and Technology Policy and a doctorate in Engineering Systems. [zszajnfa@gwu.edu]

Abstract

Complex engineered systems with long life cycles can expect to face operational uncertainty. Systems can either be flexible and change in response to a change in operating environment or can be robust and maintain system performance despite the change in operating environment. There is a wealth of literature surrounding how to design systems to be either flexible or robust, but the literature's understanding of how systems that are already in use can be modified to operate in changing circumstances is incomplete. This paper examines how numerous aircraft were modified post-production to gain new capabilities for close air support in Operation Desert Storm. Through an inductive case study, the authors find that new capabilities can be gained through changes to form and changes to tactics. Additionally, there is an interaction between form and tactical changes that has not been well defined in existing literature.

Introduction

Complex engineered systems (CES) provide critical functions for society, from power generation to mass transportation. For militaries, complex engineered systems like aircraft, missiles, and ground vehicles serve as platforms that enable crucial functions including reconnaissance, strike, and air defense. Increasingly longer and costlier development times for new systems force current systems to stay in service longer than originally planned (United States Government Accountability Office, 2019). Additionally, the constraints on logistics and budget that militaries face has led to systems taking on numerous roles, rather than having purpose-built platforms for every mission type (Wasinger, 2020). Both the need to stay in service longer and to conduct more missions with the same platform has created a need for today's military systems to be changeable.

There is a wealth of literature regarding designing systems to be changeable, but complex engineered systems may not always able to accommodate highly changeable architectures. Additionally, even with changeable product architecture, it can be difficult to plan for what must be changed. This has created a need to understand post-production change, which means changing systems after they have left production in ways not explicitly planned for during the system's design phase. We leverage a natural experiment of close air support (CAS) in Desert Storm, where aircraft were modified to perform the mission, to understand dynamics that enabled post-production change.

Literature Review

Changeability literature is focused on studying how systems can maintain value throughout their life cycle in the face of uncertain operating environments. A great deal of literature in the field is focused on how to design for changeability (Beesemyer et al., 2012; Fricke & Schulz, 2005; Ross et al., 2008). Generally, system designers choose to make flexible systems that can change in response to a change in operating environment to maintain value or to make robust systems that are able to maintain value in a variety of operating environments without having to change the system.

There are many system design properties that have been discussed such as adaptability (Li et al., 2008), scalability (Jogalekar & Woodside, 2000), and modifiability (Ross et al., 2008). These properties are meant to influence how system designers architect their systems (Fricke & Schulz, 2005; Ricci, 2014). For example, a common approach discussed is increasing the modularity of system architecture to enable changeability, but this has implications for the performance of the system as more integrated systems perform better under certain conditions (Hölttä et al., 2005).

Others in the changeability literature have focused on how to measure or quantify the changeability of systems (Rehn et al., 2019). One approach involves calculating the real age of a system after it has been modified (Enos, 2020). Studies in this field have also focused on who implements changes, be it system designers, system users, or other stakeholders (Cox, 2017).

There have been studies that do focus on the enablers of changeability. Many have focused on how margin or excess enable change (Allen et al., 2016; Eckert et al., 2019). Margin and excess refer to the system properties like power generated or weight capacity that exceed initial requirements. Since there is an excess of a property, system designers can leverage that excess to add to the system in some way. Others have focused on how change is enabled from a management perspective. One common enabler is real options in engineering design, which gives system buyers the right but not the obligation to make some change to the system in the future (de Neufville et al., 2006; Sapol & Szajnfarber, 2020).

Still, these studies are generally focused on the design phase and how system design enables change. There is limited work analyzing how complex engineered systems are modified in unexpected ways in the post-production phase (Dwyer, 2020). Long & Ferguson (2017) explore how system margin enables post-production changes to occur. Enos (2022) has taken several DoD platforms and applied his real system age formula to them to understand their useful life after being changed. Mekdeci et al. (2014) discuss how existing systems can change in response to changing contexts. This work builds on existing literature by understanding the enablers of post-production change, such as margin and modularity, and the interactions between them.

Methods

This paper leverages a case study to induce insights about how systems are modified post-production to serve new roles. The case study method is appropriate when there is not abundant empirical data to answer research questions (Eisenhardt, 2021). Cases were selected from a natural experiment where complex engineered systems, military aircraft, were changed post-production for the same mission, CAS, during the same time period, Desert Storm. Data is collected by analyzing a repository of airpower studies on Desert Storm. Insights are drawn from within and cross case examination, and these insights are then tied back to existing literature. This is a widely accepted method to induce theory from empirical case studies (Eisenhardt, 1989).

Research Setting

This paper studies how numerous aircraft were designed and modified to carry out close air support during Operation Desert Storm. Many different types of aircraft performing one mission during a relatively short war provides a natural experiment to examine how changes were made to enable CAS effectiveness, and there has been robust debate about which aircraft are best suited for CAS (Kaaoush, 2016; Macdonald & Schneider, 2016).

CAS is a core mission of air forces around the world and is defined by the U.S. Air Force as "air action by aircraft against hostile targets that are in close proximity to friendly forces and that require detailed integration of each air mission with the fire and movement of those force" (The Curtis E. LeMay Center for Doctrine Development and Education, 2020). The Marine Corps considers CAS to be a subset of its Offensive Air Support (OAS) mission set along with Deep Air Support (DAS). Requirements for OAS are taken from Marine Corps doctrine are described in the section titled "Analysis Approach."

Data Collection

Changes were identified by analyzing three documents: the DoD's official report on airpower in Desert Storm (Gulf War Air Power Survey Staff, 1993), the GAO report analyzing the claims of the DoD report (United States General Accounting Office, 1997), and a book written by the Emeritus Chair in Strategy of the Center for Strategic and International Studies that analyzes the war (Cordesman & Wagner, 2013). These documents build off and challenge each other. This is useful in being able to capture a wide spectrum of knowledge, and these documents used to triangulate information to get an accurate understanding of the changes that occurred during the war (Szajnfarber & Gralla, 2017). These documents report the instances of modifications, but other sources are used to supplement and better understand their implementation and effects where necessary.

Case Selection

The DoD identifies six aircraft mission and reports the total number of CAS sorties conducted for each: A-6 (39), A-10 (1,041), AV-8 (1,528), F/A-18 (1,978), F-16 (423), and AC-130 (31) (Gulf War Air Power Survey Staff, 1993). The A-6 and AC-130 are excluded from analysis in this study since they conducted relatively few CAS sorties, which does not provide enough data for analysis. The initial design and conception of the four analyzed aircraft are described in the Analysis section.

Analysis Approach: Functional Comparison

A functional analysis approach was taken to compare modifications against each other. Functions are derived from U.S. doctrine on CAS and serve as a way to categorize changes that occurred. The functions analyzed are effective targeting and marking, effective weaponeering, and flexible control and prompt response, which are described in this section. The unit of analysis for this paper is an attempted capability gain through a modification, be it physical or tactical. Capability gain is defined as a system gaining the ability to perform some task laid out in the definitions of the three functions that it was not able to perform in its base design.

To be able to compare how these aircraft were modified to perform CAS, it is important to understand the requirements of CAS. Marine Corps doctrine lays out requirements, which provides the baseline for our analysis: Air Superiority, Suppression of Enemy Air Defenses, Cooperative Weather, Effective Targeting, Effective Marking, Effective Weaponeering, Capable Platforms/Sensors, Flexible Control, and Prompt Response. It is important to note that these are not firm requirements that systems engineers may employ, as omission of these requirements simply means the mission is less effective, rather than mission failure (U.S. Marine Corps, 2018).

Air superiority and suppression of enemy air defenses are intended to allow CAS operations to occur without overwhelming interference from air-to-air or ground-to-air threats, respectively. They are not considered in this paper since they are conditions for a permissive operating environment rather than being a core part of the CAS mission itself. Permissive weather is a natural phenomenon that innovation and technology for CAS cannot affect, so it is not considered. The requirement for capable platforms and sensors lays out the need for technologically advanced systems that can enable mission success. Since this requirement is simply for technology to be able to make difficult aspects of the mission such as target acquisition easier, it is not considered as its own change category.

Effective targeting refers to the planning stage of a mission where commanders decide what targets will be struck, with what weapons, etc. The services and joint operations use different targeting processes, but generally, they involve assessing which targets to strike, how to strike them, striking them, and assessing the results of the strike (U.S. Marine Corps, 2018).

Effective marking is the battlefield operation of making targets visible for strike, which has been done by white phosphorus, GPS, lasers, etc. (U.S. Marine Corps, 2018). Marking is especially important for CAS as it helps reduce the likelihood of fratricide. Taken together, effective targeting and marking enable CAS operations to know what targets they are striking. This is especially important with close air support since targets are in close proximity to friendly forces.

Effective weaponeering refers to the need for “effective aircraft and weapon to target match” (U.S. Marine Corps, 2018, p. 15). Plainly, this means that weapons must be available to strike various targets and that aircraft need to be able host these weapons. This is again crucial due to the proximity of friendly forces to the targets. Weapons and the platforms delivering them need to be able to destroy the target without harming friendly forces.

Flexible control refers to the need for effective command, control, and communications (C3) to be leveraged to ensure that troops on the ground are receiving proper and timely CAS (U.S. Marine Corps, 2018). While many changes in this field revolve around communications avionics, different properties of aircraft can also enable different C3 to flow differently. For example, command structures might differ if aircraft are loitering in area to provide continuous CAS versus if they are quickly moving in and out of the battlefield.

Prompt response refers to how quickly CAS can be delivered. This is achieved by forward basing, alert states, and mission classification (U.S. Marine Corps, 2018). Forward basing moves aircraft closer to the battlefield to reduce time to battlefield. Alert states direct aircraft to be ready for takeoff within a certain amount of time and can enable faster response when on high alert. Mission states can be categorized into preplanned or on-call missions, which yield different requirements (U.S. Marine Corps, 2018). Similar to flexible control, properties of aircraft, like their speed and loiter time, may affect how mission planners are able to use different aircraft. Taken together, flexible control and prompt response mean that planners need to get CAS where it is needed in a timely manner.

Analysis

The analysis section will lay out the initial design of the aircraft identified and explain how they were meant to achieve CAS in their base configurations, based on the three functional categories identified. Then, changes that provide capability gain will be discussed. These changes are grouped by the three functional categories.

A-10 Thunderbolt

The A-10 Thunderbolt is the Air Force's first dedicated CAS platform (United States Air Force, 2019). The A-10 is often described as a plane designed around a gun, that gun being a seven barrel, 30-mm Gatling gun that is able to destroy armored ground targets (Smallwood, 1993). The aircraft was designed to be able to handle the intensity of firing the massive gun while maintaining stability and accuracy. Outside of its large cannon, the A-10 is well known for its survivability features. Many of the A-10's system requirements for survivability were derived from studying the failures of aircraft in the Vietnam War (Jacques & Strouble, 2010). Some of the survivability features include self-sealant lined fuel tanks, redundant manual controls, and a titanium casing around the cockpit (Smallwood, 1993). Ultimately, the goal of the A-10 design was to create a slow moving, highly survivable aircraft to support an enormous cannon that could destroy enemy armor with the precision of cannon fire as opposed unguided bombs or precision missiles, which the A-10 is also capable of carrying on its 11 pylons (United States Air Force, 2019).

The A-10 typically serves as a daytime attack aircraft, achieving effective targeting and marking through visual acquisition of targets, aided by aiming references and avionics such as a heads-up display. The biggest upgrade the A-10 received in targeting and marking prior to Desert Storm was the low altitude safety and targeting enhancement system, which provided constantly computed impact points for gravity bombs. Contributing to both targeting and weaponeering, the A-10 also received the Pave Penny pod, which enabled it to use laser-guided munitions (Air Combat Command, 2015). The A-10's massive cannon contributes most to its effective weaponeering capabilities, as does its ability to carry a variety of explosive munitions. System planners can use the A-10 in various ways to achieve flexible control and prompt response. The A-10 is able to loiter at low speeds, meaning it can hover in the battlefield longer than most aircraft.

AV-8B Harrier II

The AV-8B is a McDonnell Douglas (now Boeing) vertical/short take-off and landing (VSTOL) capable aircraft manufactured for the U.S. Marine Corps. It replaced the British developed AV-8, entering service with Marines in 1985 (Naval History and Heritage Command, 2014). It serves as the Marines' primary attack aircraft, and its VSTOL capabilities mean it can be deployed from the smaller carriers and austere bases that the Marines operate out of. The AV-8B was designed to perform a variety of missions like close air support but was not optimized around the role like the A-10. The focus of the Harrier design is its innovative VSTOL capabilities.

For targeting and marking, the Harrier's nose is mounted with an avionics suite that "has a TV/laser target seeker and tracker" (Gulf War Air Power Survey Staff, 1993, p. 60). The Harrier hosts a 25-mm cannon and can use traditional gravity bombs (Cordesman & Wagner, 2013). The Harrier lacked the ability to use laser-guided weapons during Desert Storm. In regard to flexible control and prompt response, the Harrier's VSTOL capabilities gives planners options to be closer to the battlefield as ships and austere runways could be setup closer to the battlefield than traditional bases. Being closer to the battlefield increases security threats but also enables quicker time to the field.

F/A-18 Hornet

The F/A-18 Hornet was manufactured for both the Navy and Marine Crops. The F/A-18's original design came out of the competition for what would become the F-16. Congress directed the Navy to consider the two prototypes from that competition, the YF-16 and YF-17, for their needs. Eventually the YF-17's revamped design, more fit for carrier operations than it had been when it was first submitted for the Air Force, won out and became the F/A-18 (Naval History and

Heritage Command, 2014). The unique designation F/A refers to the fact that it was designed to serve both air-to-air and air-to-ground missions and can be changed to either its fighter or attack configuration with external equipment (U.S. Naval Academy, n.d.).

The F/A-18s had a forward-looking infrared radar (FLIR) targeting and navigation system that enabled day and nighttime marking and targeting. The system provided thermal displays of the battlefield in real time at night, which along with night vision goggles and digital maps, enabled F/A-18s to conduct nighttime operations with largely the same procedures used for daytime operations. During the day, F/A-18s would visually acquire targets, often with the aid of binoculars, which is similar to the A-10's daytime operations (Gulf War Air Power Survey Staff, 1993). F/A-18s host a 20-mm cannon and can carry a variety of gravity and guided munitions, being one of the most versatile platforms in terms of the types of weapons it can host. This also gives planners flexibility in planning as the F/A-18 was able to engage air-to-air and air-to-ground in the same sortie (Gulf War Air Power Survey Staff, 1993).

F-16 Fighting Falcon

The F-16 came out of the Air Force Lightweight Fighter technology demonstration, with the eventual winner of the follow-on program being General Dynamics (now Lockheed Martin) (Bjorkman, 2014). The F-16 is a multi-role aircraft, designed to be able to perform air-to-ground missions while still keeping its fighter capabilities in air-to-air combat. The F-16 is fast and highly maneuverable, capable of carrying both unguided and precision weapons and still lighter than previous generation aircraft (Gulf War Air Power Survey Staff, 1993).

The F-16s deployed in Desert Storm were primarily daytime only aircraft and acquired targets visually. The F-16 hosts a 20-mm cannon and can carry gravity and guided weaponry, but the F-16 primarily used gravity and guided bombs against ground targets rather than its cannon (Gulf War Air Power Survey Staff, 1993). F-16 is a very versatile aircraft, similar to many other multi-role fighter aircraft and provides planners with flexibility similar to the F/A-18.

Changes to Enable Effective Targeting & Marking

For platforms carrying out CAS to be effective, they "need accurate weapon systems and sensor equipment to aid in target acquisition/designation in day and night operations" (U.S. Marine Corps, 2018, p. 16). A key success in Desert Storm was the ability to conduct aerial operations during nighttime. Table 1, Bombing Capabilities by Platform, modified from the DoD report, shows the system modifications that enabled nighttime visual bombing for the F-16, A-10, and F/A-18 (Gulf War Air Power Survey Staff, 1993). As described previously, the F-16 and A-10 traditionally acquire targets visually, so nighttime operations were not explicitly designed for.

Table 1: Bombing Capabilities by Platform

	Visual Bombing During Day	Visual Bombing During Night	Radar	Air-to-Air Swing Role	Comments
F-16C	X	LANTIRN-equipped aircraft	X	X	LANTIRN pods available for only two squadrons
A-10	X	X			Precision accuracy with 30-mm GAU-8 cannon; limited night capability with IIR AGM-65
F/A-18	X		X	X	Highly capable air-to-air attack aircraft

USAF General Bill Creech recognized that being able to operate at night was crucial for the United States. In 1981, Gen. Creech testified to Congress that flying at night would enable the United States to deliver more firepower 24/7, take away the cover of darkness from foes, and provide air support to ground troops engaged in nighttime combat (Slife, 2004). The Low Altitude Navigation and Targeting Infrared for Night (LANTIRN) was developed by Martin Marietta (now Lockheed Martin) to help fill this need. LANTIRN development was initiated in the 1980s and was delivered to the Air Force just before the start of the Gulf War, being installed on F-15s and F-16s (Lockheed Martin, n.d.). The system "consists of a navigation pod and a targeting pod integrated and mounted externally beneath the aircraft" (Air Combat Command, 2005). The pods can be used together or individually. The pod was designed to be very modular and was optimized around the aircraft's aerodynamic profile. It was a crucial part of the Block 40/42 upgrades that were part of a larger program to evolve from the F-16A to the F-16C (Camm, 1993). F-16s with the LANTIRN system only received the navigation pod, as the targeting pods were reserved for F-15s (United States General Accounting Office, 1997). The navigation pod uses infrared sensors to show the terrain in the heads-up display (Camm, 1993). The LANTIRN system provided the F-16 with the ability to operate at night and was crucial in denying the Iraqi army the cover of night, but there are limitations to its ability to operate in adverse weather.

While the DoD and contractors boasted about the system's performance, the GAO found that "its ability to find and designate targets through clouds, haze, smoke, dust, and humidity ranged from limited to no capability at all" (United States General Accounting Office, 1997, p. 26). Issues with targeting went unresolved throughout the war and throughout the program's duration. Despite the challenges of flying without the already insufficient targeting pods, F-16s were able to achieve nighttime bombing accuracy by flying at lower altitudes, as permitted by LANTIRN (Gulf War Air Power Survey Staff, 1993). In the case of LANTIRN, the technical change of adding a modular pod to the body of the aircraft enabled operational changes in the ability to effectively operate at night.

USAF Lieutenant Colonel Rick McDow and others realized that A-10 would likely also be called in to perform nighttime CAS since the Army was training for nighttime operations. A-10s achieved nighttime capabilities through less technically sophisticated means than the F-16. His group practiced using flares to illuminate targets on the ground, a tactic that dated back to World War II (Smallwood, 1993). Eventually, A-10 pilots realized they could fly "at night using the infrared video of the AGM-65D Maverick missile as a 'poor man's FLIR'" (Gulf War Air Power Survey Staff, 1993, p. 54). The infrared Mavericks contained a camera in the head that would feed to a small television in the cockpit (Cordesman & Wagner, 2013).

This technique was actually strongly advised against in USAF Weapons School, as looking through a camera during the daytime led to predictable flight patterns, giving air defense systems an easy target (Smallwood, 1993). The A-10 pilots, however, found relative security at night as their dark jets were hard to detect since much of the Iraqi radar capability had been crippled. Additionally, A-10 pilots had learned that their aircraft could not be heard on the ground about 5,000 feet. To maintain the cover of darkness, A-10 pilots had to fly with their lights off which made simple tasks like reading a map more difficult, and the A-10 autopilot system could not be engaged at night, taking focus away from the pilots (Smallwood, 1993). These and other challenges associated with flying at night were eventually overcome with patience and practice by the pilots. The A-10s were able to effectively deliver firepower at night and the cover of darkness they were able to utilize made flying at night extraordinarily safe (Smallwood, 1993).

Changes to Enable Effective Weaponeering

In the original conception of the F-16, it was supposed to be a lightweight, affordable air-to-air platform, a far cry from its multi-role air-to-ground capable status in Desert Storm. This was discussed in 1977 hearings on military posture, with USAF General Slay remarking that the F-16 can do CAS but is not optimized to do so like the A-10 (United States Congress House Committee on Armed Services, 1977). While the F-16 was capable but not optimized, logistic and budget concerns around retaining a CAS-oriented aircraft led the Air Force to explore the option of making the F-16 better at CAS so that the A-10 could be replaced. The initiatives to create a CAS oriented F-16 were called the A-16 or F/A-16, and these initiatives led to many technologies being test bedded (United States General Accounting Office, 1989). One technology that made its way to Desert Storm was a 30-mm gun pod marketed as GEPOD 30 and called GPU-5/A during operations. These pods were part of the Pave Claw initiative, part of the larger A-16 or F/A-16 efforts (Smith, 2021).

30-mm gun pods were attached to select Air National Guard F-16s (Gulf War Air Power Survey Staff, 1993). The gun itself was a pared-down version of the A-10's gun that was meant to be used for gun pod attachments. During the planning phase of the 30-mm pods, the goal was to provide the "flexibility to destroy heavy armor that cannot be effectively destroyed by 20 mm guns" so that the Air Force could "supplement armored vehicle killing capability of A-10" (United States Congress House Committee on Armed Services, 1980, p. 1740). In essence, the goal of the pod was to lend the A-10's unique gunpower to the F-16.

The gun pods are meant to be highly modular for the aircraft they are bolted on to as they contain their own power supply and ammunition. For the F-16, the pod was mounted on to the center of the aircraft and came with a software package to enable targeting with the pod. The modified F-16s were given to a squadron that previously flew A-10s (Werrell, 2003). During Desert Storm, the vibration from firing made the pods inaccurate and unstable. Since the pod carried such an enormous gun, the "structure of the modified F-16s could not withstand the vibrations" from firing the GPU-5/A (Smith, 2021, p. 45). While the GPU-5/A seemed like a promising way to enable the F-16 to perform CAS just like the A-10, the modification was abandoned after one day of use in Desert Storm.

Much has been said about how Desert Storm showcased the United States' capability in laser-guided and other precision weapons, including the Maverick missile, which has been discussed in relation to its usage on the A-10, the primary carrier of the Maverick during Desert Storm (Gulf War Air Power Survey Staff, 1993). The Maverick provided precision strike capability for CAS, but the extent of this capability is questionable since only 8% of the weapons fired in Desert Storm were guided (United States General Accounting Office, 1997). Additionally, all platforms that were firing Maverick missiles were able to use unguided weaponry to strike targets as well. Since Desert Storm lacked many of the close quarters combat situations that other missions like Operation Anaconda required, there was not an absolute need for guided weaponry for CAS. Mavericks merely enhance the accuracy of strikes carried out. Other platforms like the F/A-18 dropped flares and rockets to mark targets for CAS, but the use of flares and rockets to mark targets is neither innovative nor a contribution to effective weaponeering (Gulf War Air Power Survey Staff, 1993). While many new weapons were fielded, only the GPU-5/A and Maverick were identified as having contributed new capabilities for the CAS mission.

Changes to Enable Flexible Control and Prompt Response

Prior to Desert Storm , U.S. forces had been thinking about what a war in the region might look like. U.S. planners had considered that CAS would be needed if Kuwait was invaded by ground, and they wanted to make sure that there was a plan in place to provide it. USAF General Horner developed the Push CAS model April 1990 to ensure that ground forces' CAS needs were being met while not wasting resources by having aircraft be on high alert on a runway (Gulf War Air Power Survey Staff, 1993). Push CAS is a tactic that involves sending aircraft out to positions to be on-call for CAS, and allowing them to engage in DAS if CAS calls do not come in (Gulf War Air Power Survey Staff, 1993). This created a steady flow of CAS resources into the battlefield, while reducing overall inefficiency of being on alert or on station with no mission. When CAS requirements were low, all four analyzed aircraft went on to perform other missions.

No form changes were identified that enabled prompt response, but tactical decisions were made that compensated for the lack of form change in this category. Even with Push CAS in place, there was a need for CAS to be readily available in places it might not be expected. Urgent CAS calls necessitated prompt response, and the Marine Corps' Harrier was well suited for this need. The AV-8B Harrier was based closest to the Kuwaiti border and was the most forward-based aircraft of the entire war (Naval Air Systems Command, 2012). Being able to more easily shift around where the aircraft was based "allowed the aircraft to minimize its limitations in range-payload, reduce its need for refueling, and increase its sortie rate" (Cordesman & Wagner, 2013, p. 431). In this instance, military planners were able to use tactics like forward basing to overcome the AV-8B's flying limitations. It is important to note that the ability to fly the Harrier from such positions is because of its designed VSTOL capabilities.

Other changes that helped improve flexible control and prompt response did occur, but they did not provide new capabilities. Rather, many form changes in this area were improvements on existing capabilities. More capable radars and sensors made the coalition more effective but did not yield any capabilities that did not exist before. Despite the relative lack of capability gain in this area, there were many issues with communications. Coalition forces often used Soviet technology that was similar to what Iraq was fielding, so the coalition painted symbols and put lights on tanks that could be confused for Iraqi vehicles (Powell, 1991). Still, these efforts proved to marginally successful as Desert Storm saw numerous incidents of friendly fire. After the war, many tactics and technologies were developed around ensuring better communication and control of CAS missions while still ensuring prompt response time (Powell, 1991). Since these changes occurred after the war, they are out of scope of this paper.

Discussion & Conclusions

This paper analyzed how different aircraft were modified for CAS to understand what enables post-production change and how the changes are actually implemented. Changes involved both the physical modification of the aircraft and tactical innovation surrounding its use. Both of these types of changes yielded capability gains. Consistent with findings of previous authors, margin and modularity both proved useful in enabling post-production change, but not all dynamics behind post-production are well understood in the literature.

Implications for Changeability

When analyzing the changes that led to new capabilities for aircraft performing CAS, it is clear that tactical innovation can provide changeability for systems in a similar the way physical changes to form can. A change to form is not always needed to achieve the desired functional gain. For enabling nighttime operations, the F-16 and A-10 show this paradigm quite well. The F-16 was able to fly at night by equipping the LANTIRN pod, a technological advancement in aircraft navigation. The A-10 was able to achieve its nighttime capabilities first by flying in pairs and dropping flares, and then by the A-10 pilots' use of the camera on the Maverick missile. The innovation with F-16 was largely the addition of a modular pod enabling nighttime operations, while the A-10 had to use conventional tactics to achieve nighttime capabilities.

While the Maverick is technologically advanced in its own right, it required tactical innovation to be used to conduct nighttime operations. Official USAF instruction was to explicitly *not* use the Maverick's camera for navigation, but by understanding how tactical changes could be made to leverage the capability, A-10 pilots were able to challenge conventional wisdom.

This is an important insight for maintaining system life-cycle value. Functional gains can be made by innovating how systems are used. This is also shown in how flexible control and prompt response were achieved. In those cases, post-production changes to form did not occur, but military planners were able to leverage the aircrafts' inherent system properties to achieve effective CAS in ways that had not been done before the war.

The GPU-5/A is an example of why a change to form is considered risky. The system was designed to be modular, but the interfaces and system interactions were not carefully considered. This is a well understood problem in the systems engineering literature, but the GPU-5/A shows why changeable architecture is not a silver bullet for enabling post-production change. The GPU-5/A was not able to transfer the A-10's capabilities to the F-16 as intended, but the F-16 was able to perform CAS by leveraging its own strengths.

Largely, this case study reveals that margin and modularity are useful tools in enabling post-production change. Having hardpoints on aircrafts gives the system flexibility in terms of what is being changed. Specific changes were not planned when the hardpoints were installed, but they provided an interface for modules like the LANTIRN or the GPU-5/A to be integrated with the system. This is not a new concept in changeability literature, but few authors have made the connection between the physical form of the system and the operational context. This data reveals the need to look at the physical system in its operational context. Changeability literature largely focuses on the physical system, but many of the capability gains happened through tactical changes or through a combination of form change and tactical innovation. While these are not new concepts separately, they have not been well connected together. Form change and tactical change are not entirely independent from one another, as the A-10 would not be able to achieve the level of success it did at night without the Maverick's IR camera.

Future Work

This paper presents preliminary results of a larger study focused on inducing the dynamics and enablers of post-production change. There are a limited number of cases analyzed in the scope of this paper, which limits the degree of theory that can be induced from these cases. Future work will build deeper case histories in this research setting. Additionally, future work will also analyze the C-130 as a foil to this experiment. The C-130 has been modified numerous times to serve new missions, including CAS. Where this paper analyzes how numerous aircraft were modified for the one mission, the C-130 research will analyze how one aircraft was modified for numerous missions. Future work will enable more robust insights to be made

References

Air Combat Command. (2005, October). *LANTIRN*. https://www.acc.af.mil/About-Us/Fact-Sheets/Display/Article/199177/lantirn/

Air Combat Command. (2015, September). *A-10 Thunderbolt II*. 10th Air Force. https://www.10af.afrc.af.mil/About-Us/Fact-Sheets/Display/Article/966601/a-10-thunderbolt-ii/

Allen, J. D., Mattson, C. A., & Ferguson, S. M. (2016). Evaluation of system evolvability based on usable excess. *Journal of Mechanical Design*, 1–9.

Beesemyer, J., Ross, M. A., & Rhodes, D. H. (2012). An empirical investigation of system changes to frame links between design decisions and ilities. *Procedia Computer Science*, 31–38.

Bjorkman, E. (2014, March). The outrageous adolescence of the F-16. *Smithsonian Magazine*. https://www.smithsonianmag.com/air-space-magazine/outrageous-adolescence-f-16-180949491/

Camm, F. (1993). *The F-16 multinational staged improvement program.* RAND Corporation.

Cordesman, A. H., & Wagner, A. R. (2013). *Lessons of the Gulf War: 1990–1991.* Center for Strategic and International Studies.

Cox, A. M. (2017). *Functional gain and change mechanisms in post-production complex systems.* The George Washington University ProQuest Dissertations Publishing.

The Curtis E. LeMay Center for Doctrine Development and Education. (2020, October 21). *Close air support fundamentals.* https://www.doctrine.af.mil/Portals/61/documents/AFDP_3-03/3-03-D04-CLFundamentals.pdf
de Neufville, R., Scholtes, S., & Wang, T. (2006). Real options by spreadsheet: Parking garage case example. *Journal of Infrastructure Systems*, 107–111.
Dwyer, D. M. (2020). Changeability analysis for existing systems. *Australian Journal of Multi-Disciplinary Engineering*, *16*(1), 43–53.
Eckert, C., Isaksson, O., & Earl, C. (2019). Design margins: A hidden issue in industry. *Design Science*, 1–24.
Eisenhardt, K. M. (1989). Building theories from case study research. *The Academy of Management Review*, 532–550.
Eisenhardt, K. M. (2021). What is the Eisenhardt Method, really? *Strategic Organization*, 147–160.
Enos, J. R. (2020). Developing a theoretical real system age. *IEEE Systems Journal*, *15*(1), 509–517.
Enos, J. R. (2022). Applying real system age to DoD systems. *Systems Engineering*, 1–12.
Fricke, E., & Schulz, A. P. (2005). Design for Changeability (DfC): Principles to enable changes in systems throughout their entire lifecycle. *Systems Engineering*, *8*(4), 342–359.
Gulf War Air Power Survey Staff. (1993). *Gulf War air power survey.* DTIC.
Hölttä, K., Suh, E. S., & de Weck, O. (2005). Tradeoff between modularity and performance for engineered systems and products. *International Conference on Engineering Design* (pp. 1–14). The Design Society.
Jacques, D. R., & Strouble, D. D. (2010). *A-10 Thunderbolt II (Warthog) systems engineering case study.* Air Force Center for Systems Engineering, Air Force Institute of Technology.
Jogalekar, P., & Woodside, M. (2000). Evaluating the scalability of distributed systems. *IEEE Transactions on Parallel and Distributed Systems*, *11*(6), 589–603.
Kaaoush, K. J. (2016). The best aircraft for close air support in the Twenty-First Century. *Air & Space Power Journal*, 39–53.
Li, Y., Deyi, X., & Gu, P. (2008). Design for product adaptability. *Concurrent Engineering: Research and Applications*, 221–231.
Lockheed Martin. (n.d.). *A LANTIRN in the night*. https://www.lockheedmartin.com/en-us/news/features/history/lantirn.html
Long, D., & Ferguson, S. (2017). A case study of evolvability and excess on the B-52 Stratofortress and F/A-18 Hornet. *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. 1–11). American Society of Mechanical Engineers.
Macdonald, J., & Schneider, J. (2016, March 16). *Views from the ground on the A-10 debate*. War on the Rocks. https://warontherocks.com/2016/03/views-from-the-ground-on-the-a-10-debate/
Mekdeci, B., Ross, A. M., Rhodes, D. H., & Hastings, D. E. (2014). Pliability and viable systems: Maintaining value under changing conditions. *IEEE Systems Journal*, 1173–1184.
Naval Air Systems Command. (2012). *AV-8B Harrier*. https://www.navair.navy.mil/product/AV-8B-Harrier
Naval History and Heritage Command. (2014, May 29). *AV-8B Harrier II*. https://www.history.navy.mil/research/histories/naval-aviation-history/naval-aircraft/current-aircraft-inventory/av-8b-harrier-ii.html
Naval History and Heritage Command. (2014, May 29). *F/A-18 Hornet and F/A-18 Super Hornet*. https://www.history.navy.mil/research/histories/naval-aviation-history/naval-aircraft/current-aircraft-inventory/f-a-18-hornet.html
Powell, S. M. (1991, December 1). Friendly fire. *Air Force Magazine*. https://www.airforcemag.com/article/1291friendly/
Rehn, C. F., Pettersen, S. S., Garcia, J. J., Brett, P. O., Erikstad, S. O., Asbjørnslett, B. E., ... & Rhodes, D. H. (2019). Quantification of changeability level for engineering systems. *Systems Engineering, 22*(1), 80–94.
Ricci, N., Rhodes, D. H., & Ross, A. M. (2014). *Conference on Systems Engineering Research.* Redondo Beach, CA: Procedia Computer Science.
Ross, A. M., Rhodes, D. H., & Hastings, D. E. (2008). Defining changeability: Reconciling flexibility, adaptability, scalability, modifiability, and robustness for maintaining system lifecycle value. *Systems Engineering*, *11*(3), 246–261.
Sapol, S., & Szajnfarber, Z. (2020). Revisiting flexibility in design: An analysis of the impact of implementation uncertainty on the value of real options. *Journal of Mechanical Design*, 1–11.

Slife, J. C. (2004). *Creech Blue.* Air University Press.
Smallwood, W. L. (1993). *Warthog: Flying the A-10 in the Gulf War.* Brassey's.
Smith, P. C. (2021). *Fairchild Republic A-10 Thunderbolt II: The "Warthog" Ground Attack Aircraft.* Pen and Sword.
Szajnfarber, Z., & Gralla, E. (2017). Qualitative methods for engineering systems: Why we need them and how to use them. *Systems Engineering*, 497–511.
U.S. Marine Corps. (2018). *Offensive air support.* United States Department of the Navy.
U.S. Naval Academy. (n.d.). *F/A18 Hornet Strike Fighter*. https://www.usna.edu/NavalAviation/FixedWingAircraft/FA18_Hornet.php
United States Air Force. (2019, December 11). *A-10C Thunderbolt II.* https://www.af.mil/Portals/1/documents/aboutus/FactSheets/A10_Thunderbolt_II.pdf
United States Congress House Committee on Armed Services. (1977). *Hearings on Military Posture and H.R. 5068.* U.S. Government Printing Office.
United States Congress House Committee on Armed Services. (1980). *Hearings on Military Posture and H.R. 6495 (H.R. 6974).* U.S. Government Printing Office.
United States General Accounting Office. (1989). *Close air support.*
United States General Accounting Office. (1997). *Operation Desert Storm: Evaluation of the air campaign.*
United States Government Accountability Office. (2019). *Weapon systems annual assessment.*
Wasinger, M. (2020, August 17). *Airpower in the Twenty-First Century: Swing- and multi-role versus single-role specialist*. Wild Blue Yonder. https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2313184/airpower-in-the-twenty-first-century-swing-and-multi-role-versus-single-role-sp/
Werrell, K. P. (2003). *Chasing the silver bullet.* Smithsonian Books.

Correlating a User Experience System to Product Success

Jonathon M. Miller— served in the U.S. Navy as a Submariner from 2000–2006. Upon completion of military service, he attended college, whereupon he graduated with a BS in structural engineering from the University of California at San Diego in 2010. From January 2011 to present, he has worked at the Naval Information Warfare Center (NIWC) Pacific in San Diego, CA. He has performed many engineering tasks at NIWC Pacific; however, he enjoys solving complex problems which require nonlinear systems thinking. Additionally, he received an MS from the Rochester Institute of Technology in 2016 and is a PhD candidate at WPI. [Jonathon.M.Miller6.civ@us.navy.mil]

Abstract

After reviewing many products from a systems thinking perspective, a pattern has emerged in which successful products seem to more holistically address non-functional requirements (NFRs) that emphasize a user experience (UX) system. To validate this hypothesis, key NFRs – which are believed to contribute to the UX – have been identified, and pairwise comparisons will need to be made between test and control products.

To determine whether there is correlation between a UX system and success, a Spearman correlation will need to be performed in which the UX system score for each product is informed by surveys while the success of each product is informed by objective metrics.

This paper will demonstrate one subjective analysis of the independent variable and an objective analysis of the dependent variable for one pairwise comparison. Additionally, to demonstrate the mechanics of the Spearman correlation, the results of a second pairwise comparison will be included; however, its analysis will not.

Thereby, identifying the complete set of test and control comparisons and investigating whether UX systems – which holistically incorporate NFRs – and success are correlated will be the emphasis of future research and is beyond the scope of this paper.

Keywords: User Experience, Systems Engineering, Non-Functional Requirements, Systems Thinking, Requirements, Product Success

Introduction

Development teams are well-served in using the systems engineering process to develop and deploy new products. Translating stakeholder requirements into system requirements is a key step within systems engineering. Therefore, fully understanding stakeholder requirements and embedding those requirements within the design of a product is critical to success.

However, when considering stakeholder requirements, it is much more intuitive to focus on functional requirements (i.e., those requirements which technically state how a product is to perform or what a system is supposed to do) than non-functional requirements (i.e., "quality attributes or characteristics that are desired in a … [product] …, that define how a … [product] … is supposed to be" SEBoK, 2021)

Problem Formulation

With the amount of money and time that can be spent on product development efforts, every effort should be made to understand why certain efforts succeed while others do not. For example, with all of Microsoft's resources and talent, the Zune (a music player developed by Microsoft) was not nearly as successful as anticipated; however, the iPod (a music player developed by Apple) was very successful.

Why did the iPod succeed while the Zune failed?

After reviewing various products from a systems thinking perspective, a pattern has started to emerge in which successful products more holistically address non-functional requirements that emphasize a user experience system – please see Figure 1, which will be explained in more detail in subsequent sections. Also, for the purposes of this paper, a user experience system is a system that has been strategically developed to optimize the experiences that are important to the user by identifying the user as the focal point of the system in terms of the goals the user is trying to achieve and the environment in which the user is operating.

Figure 1. UX Categorization for NFRs

Therefore, the following are the objectives of this paper:

- Illustrate and categorize the envisioned user experience system.
- Describe the research methodology.
- Objectively analyze the success of a test product and a control product.
- Subjectively analyze a test product and control product against the envisioned user experience system.
- Demonstrate how to perform a Spearman correlation with the pairwise comparison that has been performed herein with a pairwise comparison that, due to the need for brevity, has not been performed herein.
- Provide a high-level description of why this research is important to the Department of Defense (DoD).
- Identify the next steps of this research.

Research Question and Hypothesis

With the problem formulation in mind, the research question is: In product development, are teams more successful when they identify their products as part of a UX system which holistically address NFRs?

Additionally, the hypothesis is: Products that are part of a UX system – which holistically address NFRs – are more successful than products that are not.

Research Objective and Contribution

If it is found that products are more successful when they are designed to be part of an overarching UX system, engineers will be better equipped in the early stages of design – especially when translating stakeholder requirements to system requirements. By generating a more robust set of requirements through a UX system perspective, there should be less churn due to updating *missed* requirements as the design progresses.

However, while better transitioning from stakeholder to system requirements is significant, the more impactful contribution is demonstrating that success is linked to the degree to which a product is part of a UX system. If it is found that success is linked to the degree to which a product is part of UX system, then it will behoove informed product development teams to holistically consider and embed NFRs into their designs.

Research Methodology

Qualitative Approach

While there are some NFRs which engineers intuitively realize are important and/or attempt to incorporate into designs due to their prevalence (e.g., RAM [reliability, availability, and maintainability]), there is not an overarching approach that facilitates the selection of NFRs needed for product development. In addition, the literature indicates that there is no agreement on a comprehensive and complete list of NFRs.

Hence, a selective literature review has been conducted in which attributes have been gleaned from successful product development efforts which seemingly provide a desirable UX system. Figure 1 illustrates the UX categorization for the NFRs – which are defined below. While this paper is limited in scope, the correlation of the degree to which designers invoke a UX system concept with product success will be attempted by performing pairwise comparisons of competing products. For each pairwise comparison, statistically relevant surveys will be utilized to assess the degree to which a UX system was used. However, with respect to product success, objective metrics will be used; this will be explained in more detail in subsequent sections.

UX Categorization: Acquisition

Acquisition is "the process of obtaining a system, product or service" (ISO/IEC/IEEE, 2017). From this UX categorization and definition, the following NFRs have been identified and defined:

- Affordability – "the degree to which the capability benefits are worth the … total life-cycle cost and support … [the user's] … goals" (Defense Acquisition University [DAU], n.d., para. 1).
- Marketability – "(of products or skills) the quality of being easy to sell because a lot of people want them" (Cambridge Dictionary, n.d., para. 1).
- Obtainability – the ability and ease with which a system or product can be obtained. This includes the removing of items from packaging.

UX Categorization: Initial Use

Here, initial use is defined as the user's first attempt to use the product. Note that, depending on packaging, setup requirements, etc., the end user may be unable to use the product on the first attempt. From this UX categorization and definition, the following NFRs have been identified and defined:

- Compatibility – "1. degree to which a product, system or component can exchange information with other products, systems or components, or perform its required functions, while sharing the same hardware or software environment 2. ability of two or more systems or components to exchange information" (ISO/IEC/IEEE, 2017, p. 79)
- Self-descriptiveness – "1. degree to which a system or component contains enough information to explain its objectives and properties" (ISO/IEC/IEEE, 2017, p. 407)

- Simplicity – "1. degree to which a system or component has a design and implementation that is straightforward and easy to understand" (ISO/IEC/IEEE, 2017, p. 414)

UX Categorization: Utilization

For utilization, it is defined as the act of the end user manipulating the product or system. From this UX categorization and definition, the following NFRs have been identified and defined:

- Operability – "1. degree to which a product or system has attributes that make it easy to operate and control" (ISO/IEC/IEEE, 2017, p. 300)
- Safety – "1. expectation that a system does not, under defined conditions, lead to a state in which human life, health, property, or the environment is endangered" (ISO/IEC/IEEE, 2017, p. 397)
- Usability – "1. extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO/IEC/IEEE, 2017, p. 492)

UX Categorization: Achieving Goals

With respect to achieving goals, it is defined as the ability for the end user to successfully meet a desired objective while or after manipulating the product in the end user's environment. From this UX categorization and definition, the following NFRs have been identified[8] and defined:

- Availability – "1. ability of a service or service component to perform its required function at an agreed instant or over an agreed period of time 2. degree to which a system or component is operational and accessible when required for use" (ISO/IEC/IEEE, 2017, p. 38)
- Interoperability – "1. degree to which two or more systems, products or components can exchange information and use the information that has been exchanged" (ISO/IEC/IEEE, 2017, p. 237)
- Reliability – "1. ability of a system or component to perform its required functions under stated conditions for a specified period of time 2. degree to which a system, product or component performs specified functions under specified conditions for a specified period of time" (ISO/IEC/IEEE, 2017, p. 375)

UX Categorization: Support

Lastly, concerning support, it is defined as the product's ability to "receive services that enable [its] continued operation" (INCOSE, 2015, p. 32).

From this UX categorization and definition, the following NFRs have been identified and defined:

---

[8] "Dependability characteristics include availability and its inherent or external influencing factors, such as availability, reliability (including fault tolerance and recoverability), security (including confidentiality and integrity), maintainability, durability, and maintenance support" (ISO/IEC/IEEE, 2017, p. 375). For the achieving goals UX categorization, it was determined that availability and reliability, in addition to interoperability, would provide better precision than embedding several NFRs into dependability.

- Maintainability – “1. ease with which a software system or component can be modified to change or add capabilities, correct faults or defects, improve performance or other attributes, or adapt to a changed environment 2. ease with which a hardware system or component can be retained in, or restored to, a state in which it can perform its required functions” (ISO/IEC/IEEE, 2017, p. 260)
- Serviceability – “Serviceability is the measure of and the set of the features that support the ease and speed of which corrective maintenance and preventive maintenance can be conducted on a system” (Radle & Bradicich, 2019, para. 5).

Use of Case Studies and Surveys

Sampling Strategy

For this effort, two sampling considerations must be made:

1. The number of pairwise comparisons to check for correlation between a UX system and success.
2. The number of participants needed to participate in the surveys which will be used to determine the degree to which a UX system is present in a particular product.

Sample Size of the Comparisons

With respect to the comparisons, the magnitude of correlation between the degree to which a product facilitates a UX system and success is of primary interest. Therefore, to determine the magnitude of correlation, a Spearman correlation will be utilized. The Spearman correlation is subsequently described in more detail, but it should be noted here that the Spearman correlation establishes how well two variables are ranked in comparison to one another for a set of given observations. For example, if two pairwise comparisons are made, then there are four observations which are comprised of two dependent variables and two independent variables. Additionally, if the products that are ranked first through fourth for the degree to which a UX system concept (i.e., the independent variable) was implemented are also ranked first through fourth for product success (i.e., the dependent variable), then there would a perfect Spearman correlation.

$$n = (n_o - b)\left(\frac{\omega_o}{\omega}\right)^2 + b$$

Equation 1. Second-Stage Sample Size Approximation for Spearman Correlation

The sample size needed to utilize the Spearman correlation has been determined based on the process proposed by Douglas Bonett and Thomas Wright (2000) – please see Equation 1. It has been found that $n = 40$, where 40 is rounded up from 39.51. Hence, with $n = 40$, 20 pairwise comparisons are needed. Detailing the process for obtaining the values for each variable in Equation 1 is beyond the scope of this paper but can be provided upon request.

Sample Size of Participants

Participants will be used with the goal of removing bias from the Spearman correlation analysis. Again, the Spearman correlation analysis has been used to investigate the correlation of the UX system and success. Therefore, to remove bias, Likert style scales will be used in which participants provide a numerical score indicating the degree to which the independent variable (i.e., the UX system) is present in a given scenario.

Additionally, while training will be provided to the participants, their training will be based on the author's interpretation of UX systems from both researching the samples and writing the surveys. Indeed, as previously stated, the author has gleaned the concept of UX systems from a selective review of successful and unsuccessful efforts while ensuring that the concept of UX systems is both rooted in systems engineering and applicable to systems thinking. Hence, the efforts to establish the UX system have necessarily been qualitative.

Thereby, while Likert type scales are quantitative tools, the underlying research is qualitative; hence, for future efforts, qualitative research will be supported by quantitative tools meant to mitigate bias via the use of participants. It has been determined that the number of participants should be the same as the number of comparisons – 40. To arrive at 40 participants, a statistical analysis has been conducted, but the explanation for this analysis is beyond the scope of this paper. Upon request, the methodology for identifying the number of participants can be provided.

Subjective Analysis

Analysis is still underway to determine the best pairwise comparisons from which 20 surveys can be created and sent to participants. Therefore, with respect to the independent variable, one example based on the author's subjective perspective is provided herein to support the contentions made – this pairwise comparison is between Apple's iPod and Microsoft's Zune. Furthermore, the next steps to validate the posited hypothesis will be provided.

However, concerning the dependent variable for the iPod and Zune, data that indicates the degree of commercial success has been compiled from the Securities and Exchange Commission (SEC) and various news reports and journals. Hence, the data pertaining to the dependent variable is objective.

With respect to the independent variable (i.e., the degree to which a UX system has been produced), a Likert style scale is used and the scoring is based upon the NFRs which have been previously defined. Therefore, for each NFR within a category, a Likert score will be obtained from zero to four; please see Table 1. Thereby, for each category, the maximum score possible is four multiplied by the number of NFRs within a particular category; the number of NFRs within the categories vary. Lastly, to obtain the overall score, the scores from each category will be summed.

Table 1. NFR Scoring

Classification	Points
Strongly disagree	0
Disagree	1
Neutral	2

iPod vs. Zune

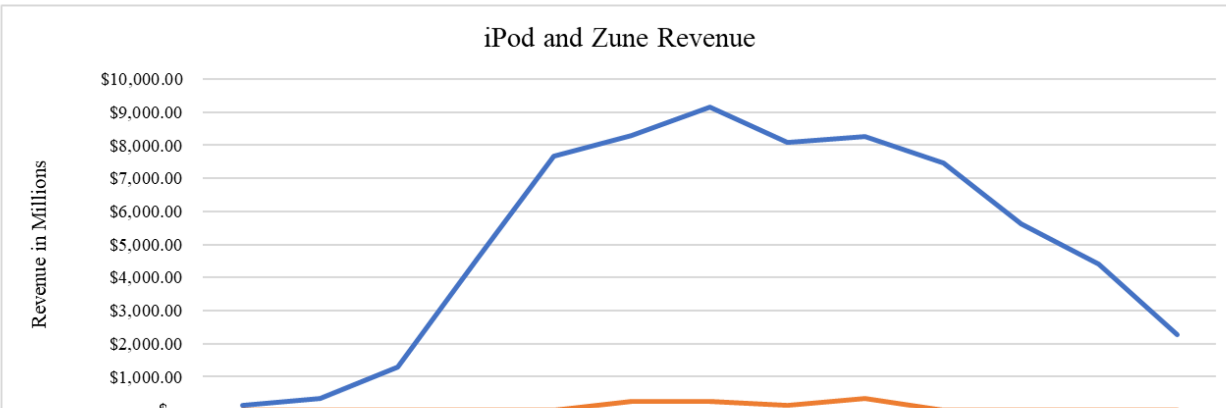
Dependent Variable

Graph 1 contains the revenue for both the iPod and the Zune. To convert the information from Graph 1 into one score for the iPod and one score for the Zune, an experimental global scale has been identified for use. "An experimental global scale is based upon an individual's personal experience with or knowledge of a particular subject" (Monat, 2009, p. 498).

$$y = y_1 + (x - x_1) * \frac{(y_2 - y_1)}{(x_2 - x_1)}$$

Equation 2. Linear Interpolation

When the revenues for the iPod and Zune are summed from Graph 1, the iPod's total revenue is approximately $67.6 billion, while the Zune's total revenue is approximately $984 million. Therefore, based on this knowledge coupled with the previously defined experimental global scale in mind, $70 billion in revenue is identified as the best value and will be awarded a score of 10, while $500 million in revenue is identified as the worst value and will be awarded a score of zero. Hence, when these figures are placed into Equation 2, where the x-values are in dollars and the y-values are success scores, the iPod receives a success score of approximately 9.65, while the Zune receives a success score of approximately 0.07.



	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
iPod Revenue	$143.00	$345.00	$1,306.	$4,540.	$7,676.	$8,305.	$9,153.	$8,091.	$8,274.	$7,453.	$5,615.	$4,411.	$2,286.
Zune Revenue	0	0	0	0	0	$250.00	$250.00	$125.00	$359.37	0	0	0	0

Graph 1. iPod[9] and Zune[10] Revenue

iPod Independent Variable

While the iPod Touch can still be purchased, the primary lifecycle of the iPod spanned from 2002–2014, with its peak years occurring during the short lifespan of the Zune, 2007–2010; please reference Graph 1. Therefore, for this comparison with the Zune, the fifth

---

[9] For the years listed, iPod revenue was obtained via the pertinent 10-K filings with the Securities and Exchange Commission listed in the references herein.

[10] Desjardins (2007) reports that Zune hits one million sales. Miller (2008) reports that Zune hits two million total sales. Microsoft's 10-Q filings for December with the Securities and Exchange Commission states, "Zune platform revenue decreased $100 million or 54% reflecting a decrease in device sales" (Microsoft Corporation, 2008, p. 31). The information from the 10-Q filing was used to extrapolate for the 2009 value. Letzing (2009) reported, "More recent data from NPD Group Inc. indicates that the Zune's already slim market share may have slipped further. NPD Group analyst Ross Rubin said in the first half of this year, Zune's share was 2%, compared to about 70% for the iPod." The iPod sold 50,312,000; this implies that the total market sold approximately 71,874,285. Therefore, for the 2010 value, it is approximated that the Zune sold 1,437,485. The cost of the Zune is assumed to be $250 throughout 2007–2010.

generation iPod Nano (16 GB), which overlapped with the lifespan of the Zune and with the iPod's peak years, will be used for this comparison – please reference Image 1 (Everyi.com, n.d.).

Image 1. iPod Nano (Fifth Generation)

Acquisition

Affordability

To reiterate, this is a subjective analysis based on the author's perspective. In future research, statistically relevant surveys will determine the score for each NFR using the Likert style scale in Table 1. Also, please note that all of the subsequent ratings utilize the Likert style scale in Table 1. Therefore, with respect to the affordability analysis, the initial retail price of the fifth generation iPod Nano was $179 (Everyi.com, n.d.). Hence, with the Zune selling for $249.99, an agree rating of three for affordability is identified.

Marketability

Between the years under consideration, 2007–2010, revenue for Apple from their full suite of iPods was $33,805,000,000; therefore, with respect to marketability, a strongly agree rating of four is identified.

Obtainability

There are no known supply chain issues for the nine different colorful versions of the iPod Nano. Therefore, when contrasted with the Zune which had limited production runs of a smaller subset of colors, the ability and ease with which a user can obtain a particular color is much greater with the iPod than the Zune.

Additionally, from observing an unpackaging video[11] of the iPod Nano, the iPod is well-protected against being damaged during transit. Also, the iPod can easily be removed from its casing by simply removing tape. Furthermore, once the case is open, the contents, with necessary paperwork, can be easily deposited into the user's hand. Hence, with respect to obtainability, a strongly agree rating of four is identified.

---

[11] A YouTuber, DetroitBORG (2009), provides a video detailing the "iPod Nano 5th Generation Unboxing" experience. The URL for this video can be found in the reference section.

Total: iPod Acquisition

The total iPod score for Acquisition is 11.

Initial Use

Compatibility

From an interview with *Time*, Steve Jobs stated, “We’re the only company that owns the whole widget – the hardware, the software, and operating system … We can take full responsibility for the user experience” (Isaacson, 2011, p. 381). Via the iTunes Store, on either a macOS or a Windows operating system, the iPod works seamlessly between the hardware and software environment – please reference Image 2 (Dudovskiy, 2021). Indeed, when a new iPod is plugged into a computer with iTunes open, iTunes recognizes the new iPod and launches the iPod setup assistant[12] – see Image 3 (Apple Inc., 2009b, p. 24). Thereby, with respect to compatibility, a strongly agree rating of four is identified.

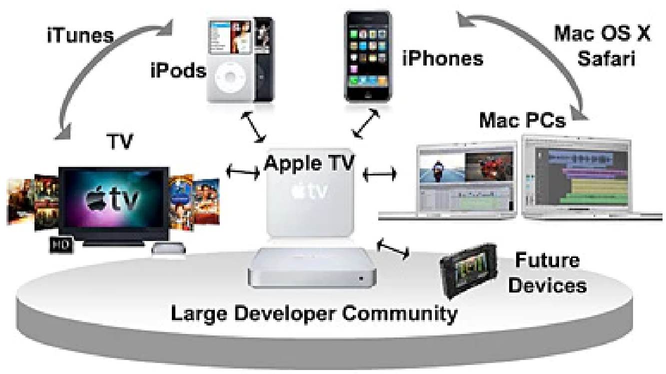


Image 2. Apple Ecosystem

Self-Descriptiveness

As seen in Image 1, the click wheel is used as the primary human machine interface (HMI). When the iPod is turned on, the user can scroll between music, videos, podcasts, etc. using the click wheel. Additionally, at the bottom of the main screen, there is a small preview panel[13] which provides a visual of the selected option (e.g., if music is selected, the album covers of the songs contained within the iPod are shown). Furthermore, with the click wheel, it is very intuitive for the user to navigate to the desired functions – see Image 4 (Apple Inc., 2009b, p. 14). Thereby, for self-descriptiveness, a strongly agree rating of four is identified.

---

[12] Fordummies (2009) has a generic video detailing the iPod setup process. The URL for this video can be found in the reference section.

[13] A YouTuber, Orangey456 (2015), provides a video which illustrates the preview pane and the ease with which the click wheel facilitates HMI. The URL for this video can be found in the reference section.

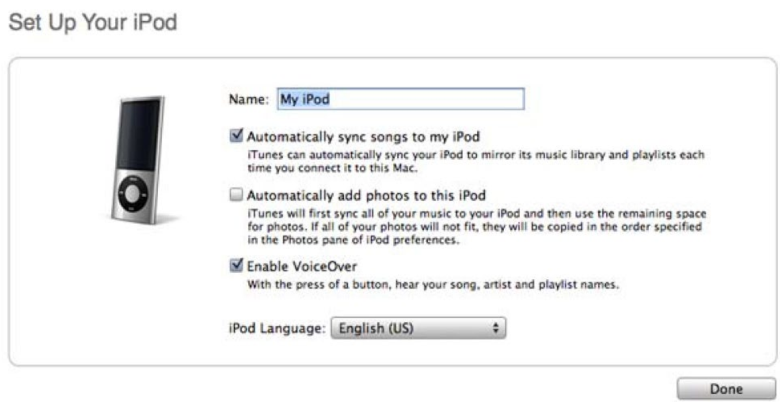


Image 3.: iTunes Setup Assistant

Simplicity

Apple realized early on that the iPod could be designed "in tandem with the iTunes software, allowing it to be simpler" (Isaacson, 2011, p. 384). Additionally, "by [Apple] owning the iTunes software and the iPod device, that allowed [Apple] to make the computer and the device work together, and it allowed [Apple] to put complexity in the right place" (Isaacson, 2011, p. 389). Hence, complexity is placed in the iTunes software (e.g., the iTunes Setup Assistant) which the computer processes, enabling the iPod to be simpler.

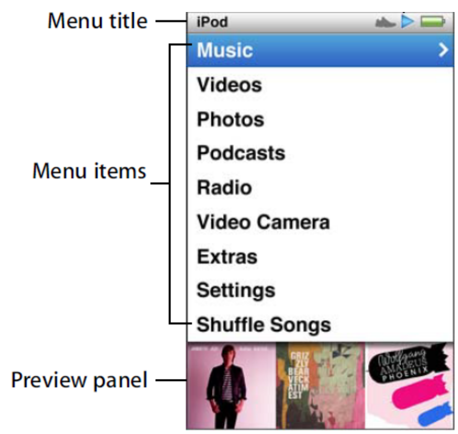


Image 4. Preview Panel with Functions

Therefore, with the integration of the iTunes software coupled with the intuitive nature of the way in which the click wheel can be used to scroll various menus, a strongly agree rating of four is identified for simplicity.

Total: iPod Initial Use

The total iPod score for Initial Use is 12.

Utilization

Operability

The degree to which the iPod can be easily operated and controlled are linked to its attributes, which includes its controls. Image 5 (Apple Inc., 2009b, p. 4) and Image 6 (Apple Inc., 2009b, p. 4) illustrate the iPod's controls and attributes.

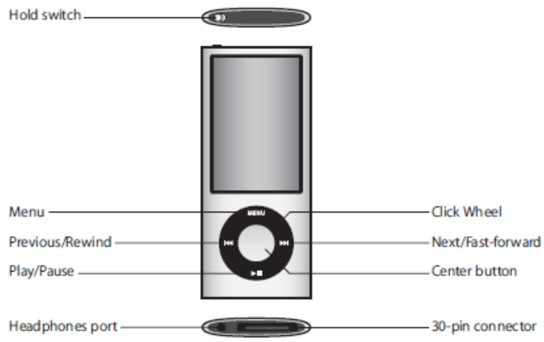


Image 5. iPod (Front View)

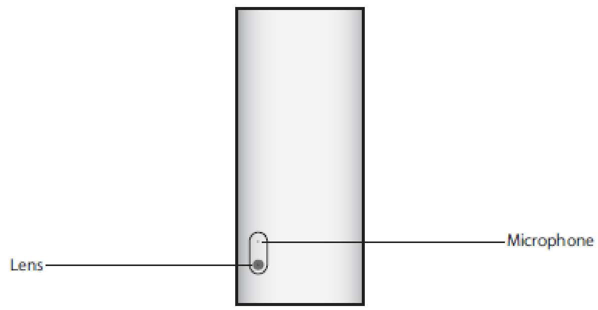


Image 6. iPod (Rear View)

As seen in Image 5, the click wheel provides the means through which the user primarily operates the iPod. As previously discussed, the click wheel provides the ability to scroll; however, it also provides intuitive "play/pause," "previous/rewind," and "next/fast-forward" functions. Hence, with the ease in which an iPod can be controlled in one hand, a strongly agree rating of four is identified for operability.

Safety

With respect to safety, the primary use of the iPod is listening to media (e.g., music, podcasts, audiobooks, etc.). Therefore, the volume on the iPod could create conditions in which the iPod is unsafe. Hence, the iPod has the ability to select a maximum volume limit. More impressively, the iPod has the ability to require a combination to change the maximum volume – which could be critical in the event a parent and/or caregiver is concerned about the judgement of the end user (Apple Inc., 2009b, p. 45). Thereby, under defined conditions, the system can help prevent endangerment. Unfortunately, there have been reports of the iPod batteries – from various models – catching fire (Marsal, 2009). However, from 2007–2009, there were approximately 160,590,000 iPods sold (Apple Inc., 2009a, p. 10). Therefore, due to the quantity of iPod sales coupled with the seemingly scarcity of incidents, the Consumer Product Safety Commission determined the risk to safety as "very low" (Marsal, 2009, para. 4). Thereby, with respect to safety, an agree rating of three is identified.

Usability

For the iPod, Jobs demanded that the user should be able to get to any song or function in three clicks or less – from any screen (Isaacson, 2011, p. 388). Therefore, in three clicks or less, the user can: listen to downloaded music, watch videos, record videos, listen to FM radio, record audio memos, etc. Hence, the user's goals with the iPod can be effectively and efficiently satisfied. Thereby, with respect to usability, a strongly agree rating of four is identified.

Total: iPod Utilization

The total iPod score for Utilization is eleven.

Achieving Goals

Availability

With respect to the iPod, mean time between failure (MTBF) and mean time to repair (MTTR) data cannot be found, and, with the exception of damage or an unusual defect, it is assumed that the iPod will be replaced by obsolescence and not due to failure. Hence, from the user's day-to-day perspective, the battery is key to availability. Therefore, with a fully charged battery, music playback is 24 hours and video playback is five hours (Apple Inc., n.d.a). Also, the battery will charge to 80% in one and a half hours and fully charge in three hours (Apple Inc., 2009b, p. 16). Thereby, with respect to availability, a strongly agree rating of four is identified.

Interoperability

"The iPod and iTunes created an ecosystem of music which included the simple purchase, organization, and portable playback that previously took several steps and suppliers to achieve" (Empringham, 2013, para. 1).

Therefore, a user's goal of listening to music became easier and more streamlined with the iPod and the interoperability of products within the Apple Ecosystem. Hence, the streamlining in Image 7 (Empringham, 2013) is made possible by the ecosystem depicted in Image 2. Thereby, with respect to interoperability, a strongly agree rating of four is identified.

Image 7. Streamlined iTunes and iPod Ecosystem

Reliability

As with availability, it is assumed that, with the exception of damage or an unusual defect, the iPod will be replaced by obsolescence and not due to failure. Again, with respect to the iPod, MTBF and MTTR data cannot be found. Hence, from the user's perspective, the battery is key to reliability – the battery has a limited number of charging cycles.

Links to information pertaining to the iPod Nano's battery Apple no longer appear to be active. Therefore, since Apple uses a lithium-ion battery for both the iPhone and the iPod, it is assumed here that the iPod's battery will behave similarly to the iPhone which "is designed to retain up to 80% of its original capacity at 500 complete charge cycles when operating under normal conditions" (Apple Inc., n.d.b). Also, the iPod has a built-in energy saver which helps decrease the rate of battery consumption (Apple Inc., 2009b, p. 18). Thereby, with respect to reliability, a strongly agree rating of four is identified.

Total: iPod Achieving Goals

The total iPod score for Achieving Goals is 12.

Support

Maintainability

With respect to the iPod, it has been previously discussed that complexity resides in iTunes on a computer in lieu of the iPod itself. Therefore, by having the most up-to-date iTunes software installed on a computer with which the iPod is synchronized, software maintenance is simple. Concerning hardware maintainability, the iPod can be easily damaged (especially the glass display) when dropped and when liquid (e.g., sweat, etc.) gets inside the casing. However, there are a plethora of cases which can be placed around the iPod which drastically minimizes the impact of getting wet (see Image 8; Amazon.com, Inc., n.d.a) and/or being dropped (see Image 9; Amazon.com, Inc., n.d.b). Thereby, with respect to maintainability, a strongly agree rating of four is identified.

Image 8. iPod Nano Armband

Serviceability

Most problems can be solved quickly by the user following “the five Rs” which is to: 1) reset the iPod, 2) retry connecting the iPod with a different USB port, 3) restart the computer and ensure the computer has the latest iTunes software, 4) reinstall the latest iTunes software, and 5) restore the iPod with the most up-to-date iTunes software (Apple Inc., 2009b, pp. 86–92).

Image 9. iPod Nano Silicone Case

As discussed during reliability, the battery will have a limited number of charging cycles; therefore, it is the most likely item needing service. Unfortunately, for the fifth generation iPod Nano, expert battery installation is required – please see Image 10 (iPodBatteryDepot, n.d.). Fortunately, Apple has extensive support services via their Genius Bar (Apple Inc., n.d.c), and the user will be charged for support based on warranty coverage (Apple Inc., n.d.d). Thereby, with respect to serviceability, an agree rating of three is identified.

Total: iPod Support

The total iPod score for Support is seven.

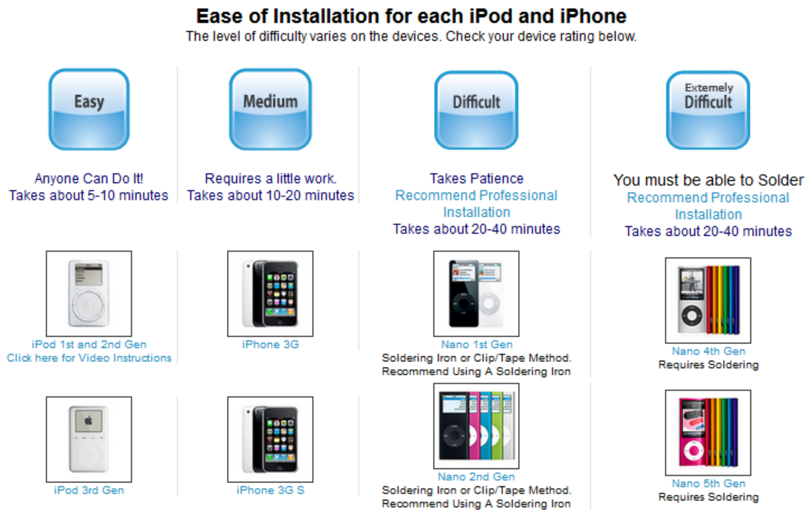


Image 10. Ease of Battery Installation for Each iPod and iPhone (Abridged)

Zune Independent Variable

Per Graph 1, the Zune generated revenue from 2007 to 2011. Additionally, Microsoft launched the Zune when iPod sales were at their peak. Therefore, for this comparison, the UX system of the initial Zune (Zune 30) offering is evaluated – please reference Image 11 (Stars and Décor Co., n.d.).

Image 11. Zune 30

Acquisition

Affordability

The initial retail price of the Zune was $249.99 (Montalbano, 2006, para. 1). Therefore, with the iPod selling at $179, a neutral rating of two for affordability is identified.

Marketability

With Apple dominating the portable media player (PMP) market in 2006 with a total revenue of $7,676,000,000 (Apple Inc., 2006, p. 54) and having a four-year head start in building up a customer base, it is difficult to envision a scenario in which the Zune would be easy to sell to anyone other than people who do not like the Apple brand and/or people who are technology enthusiasts. Hence, a strongly disagree rating of zero is identified for marketability.

Obtainability

No major supply chain concerns can be found for the Zune itself; however, there were several limited editions. For example, only 100,000 pink Zunes were produced (Quilty-Harper, 2007, para. 1). Initially, when a customer ordered a Zune, the customer had to pick between black, brown, or white. However, during the initial offerings of new and oftentimes limited-edition colors of the Zune, Microsoft randomly sent Zunes which were not black, brown, or white to customers. For example, had someone ordered one of the original colors of black, brown, or white, they might have been sent an orange Zune (Oiaga, 2006).

Additionally, from observing an unpackaging video[14] of the Zune 30 performed recently by a collector, the Zune is well-protected against being damaged during transit; however, it requires a boxcutter to efficiently unpackage the Zune. Also, various items appear to be hidden away at first glance – making them susceptible to be unseen and thrown out.

Due to these concerns, a disagree rating of one for obtainability is identified.

Total: Zune Acquisition

The total Zune score for Acquisition is three.

Initial Use

Compatibility

To initially use[15] the Zune, it must first be plugged into a Windows based operating system (OS); and, with the Zune plugged in, the installation compact disc (CD) must be inserted into the computer with an internet connection. An internet connection is needed to ensure the installation software obtains any required updates. Once the installation software has completed its tasking and the Zune is fully charged, the computer will prompt the user that the Zune is ready to be energized. Furthermore, the Zune itself may require updates – which, if needed, will begin shortly after being energized.

Hence, a disagree rating of one is identified for the Zune's compatibility due to the following: 1) the need for a CD, 2) the requirement of a Windows based OS, 3) the potential for updates on both the computer software and Zune itself, and 4) the need to wait for the computer to prompt the user to energize the Zune.

Self-Descriptiveness

Similar to the iPod, the Zune has a click wheel which allows the user to navigate between functions. In addition, per Image 11, the functions are clearly labeled. However, the Zune does not have a small preview panel – like the iPod. The Zune's albums category does

[14] A YouTuber, Lowkey Bre (2021), provides a video detailing the "Zune unboxing" experience. The URL for this video can be found in the reference section.

[15] The YouTube channel Gadgets and Gears (2011), which is a child channel to the YouTube Videojug channel, provides a video detailing the initial Zune installation and charging. This video is the basis for this section. The URL for this video can be found in the reference section.

provide similar information to the iPod, but not as a preview – please see Image 12 (Block, 2006). Therefore, an agree rating of three for self-descriptiveness is identified.

Image 12. Albums Category

Simplicity

As previously mentioned,[16] a CD and a Windows based computer – which is connected to the internet – is needed to initially use the Zune. Additionally, the user is advised to not energize the Zune until prompted by the installation software; and, after being energized, the Zune itself may require updates before its initial use.

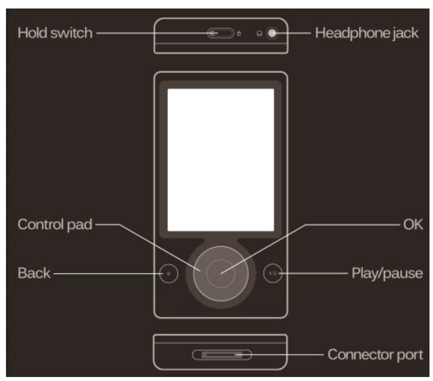


Image 13. Zune's Controls

Due to the number of specific instructions and the need for a CD, a disagree rating of one for simplicity is identified.

Total: Zune Initial Use

The total Zune score for Initial Use is five.

---

[16] The YouTube Channel Gadgets and Gears (2011), which is a child channel to the YouTube Videojug channel, provides a video detailing the initial Zune installation and charging. The URL for this video can be found in the reference section.

Utilization

Operability

The ease with which the Zune can be operated is linked to its controls – as seen in Image 13 (Amazon.com, Inc., n.d.c). Also, the controls for the Zune can be manipulated to obtain different outcomes – please see Image 14 (Microsoft Corporation, n.d.). For example, the back arrow can either allow the user to return to the previous screen or return to the home screen – depending on the duration that the button is held.

CONTROLS

On/off	Press and hold .
Volume	Press the Zune pad up or down.
Next/previous	To skip, press the right side of the Zune pad. To go back, press the left side.
Fast-forward/ rewind	To fast-forward, press and hold the right side of the Zune pad. To go back, press and hold the left side.
Playback options	For shuffle, repeat, rate, send, and song list, press the Zune pad from the Now Playing screen.
Previous screen	Press .
Return to home	Press and hold .
Lock/unlock	To lock the controls, set the lock ( ) to pink.

Image 14. Zune Controls

While the buttons appear to be simple and in an intuitive location, utilizing the multiuse buttons does not appear to be immediately intuitive; therefore, it seems that this may – at least initially – impede the ease with which the Zune can be operated. Hence, a neutral rating of two for operability is identified.

Safety

Similar to the iPod, the Zune's primary use is listening to and watching media (e.g., music, videos, audiobooks, etc.). Unlike the iPod, however, the Zune does not appear to have the ability to set a maximum audio limit – which can be critical in the event that a parent and/or caregiver is concerned about the judgement of the end user.

Additionally, while there were limited production runs, no major safety concerns – pertaining to the Zune or any of its components – can be found. Hence, a neutral rating of two for safety is identified.

Usability

Image 11 illustrates the six main areas available, and Image 12 illustrates the twist interface. As shown in Image 12, the twist interface facilitates quick access (Block, 2006, para. 10) to some of the contents of the six main areas by enabling both horizontal and vertical scrolling. While the twist interface allows the user to quickly navigate, it is operated by the control pad and seems to require a learning curve for a user to become adept at knowing what control pad manipulation operates which portion of the twist interface. Hence, an agree rating of three for usability is identified.

Total: Zune Utilization

The total Zune score for Utilization is seven.

Achieving Goals

Availability

Similar to the iPod, the Zune's MTBF and MTTR data cannot be found. In addition, as with the iPod, it is assumed that the Zune – with the exception of damage or an unusual defect – will be primarily replaced by obsolescence and not due to failure. Therefore, from a user's perspective, the battery is key to availability. Hence, with a fully charged battery, music playback is fourteen hours (Kim, 2006, para. 1) and video playback is four hours (Kim, 2006, para. 2). Furthermore, upon complete draining, the battery will fully charge in approximately three hours (King, n.d., para. 2).

Dissimilar to the iPod, all Zune 30s suffered a major outage – just as Microsoft was trying to capture market share in the PMP market. Due to the leap year in 2008, all Zunes froze on the last day of the year because they were programmed to have 365 days in lieu of 366 days. To get the Zunes to function, Microsoft advised users to allow the Zune's battery to fully drain and reenergize the devices on January 1, 2009 (Wortham, 2008).

Due to the leap year issue and a suboptimal battery when compared to the iPod, a disagree rating of one for availability is identified.

Interoperability

Similar to the way in which the iTunes Store and the iPod operated, the Zune had its own way to purchase music and other media content via the Zune Marketplace – the Zune Marketplace is no longer operational. Additionally, with the Zune Music Pass, Microsoft provided subscribers with unlimited access to the entire catalog on the Zune Marketplace. Whereas the iTunes Store also provided the primary software functions for the iPod, the Zune Marketplace and the Zune software – which is found on the installation CD – are separate. Hence, there was a series of installation steps (as previously discussed) which needed to occur before the Zune could access the Zune Marketplace. Thereby, with respect to interoperability, a neutral rating of two is identified.

Reliability

From a user's perspective, it is assumed that the battery is key to reliability. As previously stated, it is likely that the Zune will be replaced by obsolescence and not due to failure; unfortunately, there is neither MTBF nor MTTR data to verify this assumption against.

Image 15. Zune Enclosed Within a ToughSkin Case

While test data pertaining to the Zune's battery cannot be found, it is known that the battery is a lithium-ion, 3.7 volts (IFIXIT, n.d.b, para. 5). For this type of lithium-ion battery, it has been assumed that the expected life cycle is greater than 500[17] which is similar to the assumptions made about the iPod. However, unlike the iPod, the Zune does not have a built-in energy saver to help decrease battery consumption. Thereby, with respect to reliability, an agree rating of three is identified.

Total: Zune Achieving Goals

The total Zune score for Achieving Goals is six.

Support

Maintainability

While software support for the Zune is no longer available, the Zune did allow for automatic updates. To facilitate automatic updates, the Zune needed to be connected to a computer with the Zune software running – the software was designed to automatically look for new versions (IFIXIT, n.d.a, para. 5). With respect to hardware maintainability, the Zune had several case options available which could minimize the impact of being dropped or getting wet. For example, Speck made a ToughSkin case for the Zune. The ToughSkin case was highly ruggedized – please see Image 15 (Amazon.com, Inc., n.d.d).

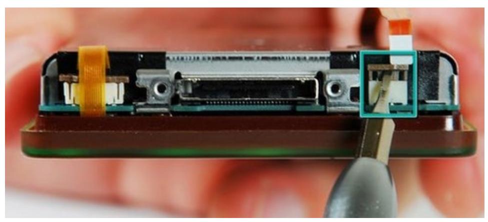


Image 16: Zune Battery Release

Due to the limited production run of the Zune which ultimately caused the ceasing of software support, it is obvious that long-term maintainability for the end user is, at best, difficult. However, if the Zune would have had greater success, leading to the production of more Zunes, it is clear that the Zune was well-positioned to provide maintainability. Thereby, due to the impact that limited production of the Zune had on maintainability, a neutral rating of two is identified.

Serviceability

Regarding the ease and speed with which corrective action can be taken on the Zune, nothing can be found in the official Zune literature describing how most problems can be resolved. It was previously conveyed that the battery is key to reliability; therefore, the battery is likely to be the primary item within the Zune needing service. Due to the way in which the battery is internally attached and the tools needed to extract the battery, replacing the battery is beyond the scope of what most users will be able to accomplish and will likely require an expert

---

[17] "Lithium Polymer Ion batteries provide the performance of the Li-ion in a thin or moldable package" (HiMAX, n.d., para. 1). "Expected cycle life is about 500+ cycles" (HiMAX, n.d., para. 3).

technician – please see Image 16 (IFIXIT, n.d.b) which illustrates the release of the battery ribbon and Image 17 (IFIXIT, n.d.b) which depicts the battery removed from the Zune.

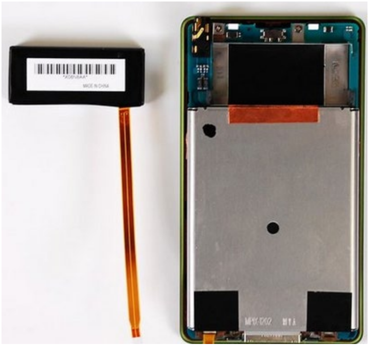


Image 17. Zune Battery Removed

Even before the Zune was discontinued, there was nothing similar to the Genius Bar that Apple provides. Hence, if the user was able to obtain service via a warranty, the user would be required to package and mail the Zune. Furthermore, as Microsoft was phasing out the Zune, they abruptly cancelled their extended warranty program (Hunting, 2011). Thereby, with respect to serviceability, a strongly disagree rating of zero is identified.

Total: Zune Support

The total Zune score for Support is two.

Independent Variable Summary: iPod and Zune

Table 2. Independent Variable Summary of the iPod and Zune

Product	Acquisition	Initial Use	Utilization	Achieving Goals	Support	Total
iPod	11	12	11	12	7	53
Zune	3	5	7	6	2	23

Subjective Spearman Correlation Example

To illustrate how the hypothesis stated herein – which asserts that product development teams are more successful when they identify their products as part of a UX system by holistically addressing NFRs – will ultimately be tested, two pairwise comparisons will be used. However, for the sake of brevity, the analysis of the independent variable and dependent variable for another pairwise comparison will not be performed herein.

$$r_s = 1 - \frac{6\sum_i d_i^2}{n(n^2-1)}$$

Equation 3. Spearman Correlation Equation

The Spearman Correlation Equation can be found in Equation 3, where “r_s” is bounded between negative one and one. The closer “r_s” is to one, the more positive the correlation – this would confirm the author’s hypothesis. However, the closer “r_s” is to negative one, the more negative the correlation. If “r_s” equals zero, there is no discernable correlation.

In Equation 3, “n” is the number of each observation (i.e., twice the number of pairwise comparisons) and “d” is equal to the distance between the success (i.e., the dependent variable) score rank and the UX (i.e., the independent variable) score rank – each observation will have its dependent and independent variable ranked. After the ranking is performed, “d” will be squared for each observation and the sum of “d^2” will be tallied. With the results from the previously performed iPod and Zune analysis, this is illustrated in Table 3 with an additional pairwise comparison of the iPhone and BlackBerry which, as previously stated, is not conducted herein for the sake of brevity.

Table 3. Validation Example

-	UX Score	Success (Global) Score	UX Score (rank)	Success Score (rank)	d	d^2
iPod	53	9.65	2	2	0	0
Zune	23	0.07	3	4	-1	1
iPhone	55	9.85	1	1	0	0
BlackBerry	20	0.65	4	3	1	1
$\sum_i d_i^2$	-	-	-	-	-	2

When the results from Table 3 are placed into Equation 3, the results are Calculation 1 which demonstrates a positive Spearman Correlation of 0.8 (i.e., the more that a product holistically addresses NFRs which emphasize a UX system, the more successful the product will be).

$$r_s = 1 - \frac{(6)(2)}{4(4^2 - 1)} = 0.8$$

Calculation 1. Spearman Correlation Calculation

Potential Importance to the DoD and Next Steps

DoD Importance

The current level of hostility around the globe has not been experienced in over 70 years. Additionally, while threats are seemingly increasing, funding to the DoD is decreasing – primarily due to increased fuel costs and inflation.

> President Biden’s $715 billion FY22 budget for the Department of Defense … assumed 2.2% inflation and a 10.1% increase in fuel costs. With inflation … closer to 7% and with fuel costs for the world’s largest purchaser of fuel up 30%, DoD faces a $40 billion second half hole in its FY22 budget. For perspective, the FY13 Budget Control Act (BCA) DoD sequester totaled $37.2 billion. (Taylor & Goss, 2022, para. 1)

Therefore, now is the time for engineering ingenuity within the DoD by thinking in terms of the warfighters’ experiences and building systems which efficiently maximize those experiences. For example:

- Placing cost in the forefront, how efficiently can a system – with the attributes important to the warfighter – be acquired?

- Pertaining to training, how can the learning curve needed to operate a warfighting system be minimized without compromising effectiveness?
- When the warfighter is under extreme pressure, how can it be ensured that a warfighting system operates in a usable and safe manner?
- With the number of interconnected systems within the DoD, how can it be ensured that a given system will be available to reliably interoperate with other systems to achieve the warfighter's goals?
- Concerning return on investment, how can warfighting systems achieve longevity via proper support?

These are just few questions[18] which are believed to be natural extensions of this research.

Next Steps

The immediate next steps in this research are to: 1) fully identify the statistically relevant set of commercial pairwise comparisons, 2) create a survey to determine the degree to which a UX system has been implemented in each observation, 3) obtain objective success data for each observation, 4) provide the surveys to current graduate students enrolled in a systems thinking class, and 5) evaluate the survey effectiveness.

Regarding the intermediate steps of this research, the following is planned: 1) the surveys will be modified based on any lessons learned from the initial surveys which will solely be sent to graduate students enrolled in a systems thinking class, 2) the surveys will be sent to the statistically relevant set of participants, 3) the data will be processed and conclusions will be made, and 4) a dissertation will be completed.

In the long-term, if it is ultimately found that commercial products which are part of a UX system – as described herein – correlate to success, then translating this insight into DoD applicability will be an emphasis of future research.

References

Amazon.com, Inc. (n.d.a). *Tune belt armband for iPod nano 4th and 5th generation - for use with or without Nike+ receiver*. Retrieved February 27, 2022, from https://www.amazon.com/Armband-Compatible-iPod-nano-generation/dp/B002RX7RN6

Amazon.com, Inc. (n.d.b). *Silicone case for iPod nano 5th generation, clear, replacement part from complete TuneBand package, silicone CASE ONLY*. Retrieved February 27, 2022, from https://www.amazon.com/Silicone-Generation-Replacement-Complete-TuneBand/dp/B01H4EVI4Q

Amazon.com, Inc. (n.d.c). *Zune 30 GB digital media player (pink)*. Retrieved March 19, 2022, from https://www.amazon.com/Zune-Digital-Media-Player-Pink/dp/B000OK0CIQ

Amazon.com, Inc. (n.d.d). *Speck ToughSkin case for Zune 30 GB (black)*. Retrieved March 19, 2022, from https://www.amazon.com/dp/B000JLU1JU

Apple Inc. (2003). *Form 10-K*. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/320193/000104746903041604/a2124888z10-k.htm

Apple Inc. (2004). *Form 10-K*. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/320193/000104746904035975/a2147337z10-k.htm

Apple Inc. (2005). *Form 10-K*. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/320193/000110465905058421/a05-20674_110k.htm

[18] It is understood and appreciated that a lot of great work is currently underway to address these questions. It is respectfully suggested that holistically addressing these questions from the perspective of Figure 1 could be beneficial.

Apple Inc. (2006). *Form 10-K*. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/320193/000110465906084288/a06-25759_210k.htm
Apple Inc. (2007). *Form 10-K*. Securities and Exchange Commission https://www.sec.gov/Archives/edgar/data/320193/000104746907009340/a2181030z10-k.htm
Apple Inc. (2008). *Form 10-K*. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/320193/000119312508224958/d10k.htm
Apple Inc. (2009a). *Form 10-K*. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/320193/000119312510012091/d10ka.htm
Apple Inc. (2009b). *iPod nano user guide*. https://www.manualsdir.com/manuals/546997/apple-ipod-nano-5th-generation.html?download
Apple Inc. (2010). *Form 10-K*. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/320193/000119312510238044/d10k.htm
Apple Inc. (2011). *Form 10-K*. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/320193/000119312511282113/d220209d10k.htm
Apple Inc. (2012). *Form 10-K*. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/320193/000119312512444068/d411355d10k.htm
Apple Inc. (2013). *Form 10-K*. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/320193/000119312513416534/d590790d10k.htm
Apple Inc. (2014). *Form 10-K*. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/320193/000119312514383437/d783162d10k.htm
Apple Inc. (n.d.a). *iPod nano (5th generation) - technical specifications*. https://support.apple.com/kb/sp569?locale=en_US
Apple Inc. (n.d.b). *iPhone battery and performance*. https://support.apple.com/en-us/HT208387
Apple Inc. (n.d.c). *For getting the right support, here's a genius idea*. https://www.apple.com/retail/geniusbar/
Apple Inc. (n.d.d). *iPod service and repair*. https://support.apple.com/ipod/repair/service
Block, R. (2006). *Zune review*. engadget. https://www.engadget.com/2006-11-15-zune-review.html
Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika*, *65*(1), 23–28.
Cambridge Dictionary. (n.d.). *Marketability*. Cambridge University Press. Retrieved February 19, 2022, from https://dictionary.cambridge.org/us/dictionary/english/marketability
Defense Acquisition University. (n.d.). *Affordability – SE trade-off analyses*. https://www.dau.edu/tools/se-brainbook/Pages/Design%20Considerations/Affordability.aspx
Desjardins, D. (2007). Zune sales hit 1 million; has ways to catch up. *Retailing Today*, *46*(10), 19.
DetroitBORG. (2009, September 13). *iPod nano 5th generation unboxing* [Video]. YouTube. https://www.youtube.com/watch?v=I6r_nP8UR28
Dudovskiy, J. (2021). *Apple ecosystem: Closed and effective*. Business Research Methodology. https://research-methodology.net/apple-ecosystem-closed-effective/
Empringham, T. (2013). *Connecting the value chain*. SucceedSooner. https://succeedsooner.ca/2013/04/08/connecting-the-value-chain/
Everyi.com. (n.d.). *Apple iPod nano 5th gen/camera 8 GB, 16 GB specs*. https://everymac.com/systems/apple/ipod/specs/ipod-5th-generation-5g-nano-specs.html
fordummies. (2009, February 25). *How to set up your iPod for dummies* [Video]. YouTube. https://www.youtube.com/watch?v=60E4WV1efYg
Gadgets and Gears. (2011, April 6). *A guide to installing and charge your Zune* [Video]. YouTube. https://www.youtube.com/watch?v=Daa3cgwLxBQ
HiMAX. (n.d.). *Li polymer 3.7V 800mAh battery*. https://himaxelectronics.com/product-item/li-polymer-3-7v-800mah-battery/
Hunting, B. (2011). *Microsoft ends extended warranty coverage for the Zune*. https://benjaminhunting.com/2011/01/microsoft-ends-extended-warranty-coverage-for-the-zune/
IFIXIT. (n.d.a). *Microsoft Zune 30 troubleshooting*. https://www.ifixit.com/Wiki/Microsoft_Zune_30
IFIXIT. (n.d.b). *Microsoft Zune 30 GB battery replacement*. https://www.ifixit.com/Guide/Microsoft+Zune+30+GB+Battery+Replacement/1230
INCOSE. (2015). *Systems engineering handbook: A guide for systems life cycle processes and activities* (4th ed., D. D. Walden, G. J. Roedler, K. J. Forsberg, R. D. Hamelin, & T. Shortell, Eds.). John Wiley & Sons, Inc.

iPodBatteryDepot. (n.d.). *Ease of installation for each iPod and iPhone*. https://www.ipodbatterydepot.com/easytoinstall.asp
Isaacson, W. (2011). *Steve Jobs* (1st ed.). Simon & Schuster.
ISO/IEC/IEEE. (2017). *ISO/IEC/IEEE 24765:2017(E): ISO/IEC/IEEE international standard - Systems and software engineering–Vocabulary*. IEEE.
Kim, J. (2006). *Inside details about Zune's battery life. Some inside details about Zune's battery life*. c|net. https://www.cnet.com/culture/inside-details-about-zunes-battery-life/
King, M. (n.d.). *How do I turn on my Zune?* CHRON. https://smallbusiness.chron.com/turn-zune-53456.html
Letzing, J. (2009). *What's to become of Microsoft's answer to the iPod?* MarketWatch. Available at: https://www.marketwatch.com/story/microsofts-zune-continues-to-struggle-2009-07-29
Lowkey Bre. (2021, October 11). *Zune unboxing in 2021| Microsoft Zune 30GB review | I bought a Zune in 2021 | Failed review in 2021?* [Video]. YouTube. https://www.youtube.com/watch?v=UO41UhPXmOA
Marsal, K. (2009). *Apple accused of stalling iPod battery fire investigation*. Apple Insider. https://appleinsider.com/articles/09/07/22/apple_accused_of_stalling_ipod_battery_fire_investigation
Microsoft Corporation. (2008). *Form 10-Q*. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/789019/000119312509009386/d10q.htm
Microsoft Corporation. (n.d.). *Zune player product manual*. https://www.microsoft.com/en-us/download/details.aspx?id=30468
Miller, P. (2008). *Zune breaks 2 million sold, stealing market from* Cr*eative not Apple*. engadget. https://www.engadget.com/2008-05-10-zune-breaks-2-million-sold-stealing-market-from-creative-not-ap.html
Monat, J. (2009). The benefits of global scaling in multi-criteria decision analysis. *Judgment and Decision Making*, *4*(6), 492.
Montalbano, E. (2006). *Microsoft's Zune to ship Nov. 14, cost $249*. Computerworld. https://www.computerworld.com/article/2547041/microsoft-s-zune-to-ship-nov--14--cost--249.html
orangey456. (2015, July 23). *iPod nano 5th generation retro review!* [Video]. YouTube. https://www.youtube.com/watch?v=jWgqIYfnLoA
Oiaga, M. (2006). *$750 orange Zune goes unsold first pink now orange*. Softpedia News. https://news.softpedia.com/news/750-Orange-Zune-Goes-Unsold-43174.shtml
Quilty-Harper, C. (2007). *Official pink Zune gets 100,000 unit production run*. engadget. https://www.engadget.com/2007-04-07-official-pink-zune-gets-100-000-unit-production-run.html
Radle, B., & Bradicich, T. (2019). *What is serviceability?* National Instruments. https://www.ni.com/en-us/innovations/white-papers/13/what-is-serviceability-.html
Stars and Décor Co. (n.d.). *Zune 30 GB digital media player (black)*. https://starsanddecor.com/products/zune-30-gb-digital-media-player-black?_pos=1&_psq=zune&_ss=e&_v=1.0 [Accessed: 19-February-2022]
Taylor, D., & Goss, C. (2022). *DoD's inflation challenge*. Federal Budget IQ. https://federalbudgetiq.com/insights/dods-inflation-challenge/
Wortham, J. (2008). The day the Microsoft Zunes stood still. *The New York Times*. https://web.archive.org/web/20130522085019/http://bits.blogs.nytimes.com/2008/12/31/the-day-microsoft-zunes-stood-still/?hp

PANEL 17. DESIGNING AND DEPLOYING ARTIFICIAL INTELLIGENCE TO IMPROVE PERFORMANCE

Thursday, May 12, 2022	
9:40 a.m. – 10:55 a.m.	**Chair: Rear Admiral William "Scott" Dillon,** Commander, Naval Air Warfare Center, Weapons Division/ Assistant Commander for Test and Evaluation, Naval Air Systems Command ***Cybersecurity, Artificial Intelligence, and Risk Management: Understanding Their Implementation in Military Systems Acquisitions*** Johnathan Mun, Naval Postgraduate School Thomas Housel, Naval Postgraduate School ***Two Gaps That Need to be Filled in Order to Trust AI in Complex Battle Scenarios*** Bruce Nagy, NAVAIR

Rear Admiral William "Scott" Dillon—is a native of Pittsburgh, Pennsylvania.He graduated from the U.S. Naval Academy in 1991, is a graduate of U.S. Naval Test Pilot School Class 117 and holds a Master of Business Administration from the University of Chicago and a Master of Science in Aeronautical Engineering from the Naval Postgraduate School.

After designation as a naval aviator in 1994, he reported to Patrol Squadron (VP) 1 at Barber's Point, Hawaii. During this tour, he completed deployments as a P-3 tactical coordinator and mission commander. Dillon's follow-on assignment was to the U.S. Naval Test Pilot School/Naval Postgraduate School Cooperative Program and subsequently to Air Test and Evaluation Squadron (VX) 20. He later deployed to Djibouti, Africa during Operation Enduring Freedom.

Dillon was selected as an aerospace engineering duty officer (AEDO) in 2002 and served in a series of acquisition billets including P-3 class desk and integrated product team lead at Maritime Patrol and Reconnaissance Aircraft Program office (PMA) 290 and MH-60R team lead at PMA-299. His acquisition tours also included assignment as the AEDO community detailer at Navy Personnel Command in Millington, Tennessee.

From May 2012 through March 2016, Dillon served as program manager, PMA-290, responsible for all P-8A and P-3 programs. During his tenure, the P-8A completed initial operational testing and evaluation, initial operational capability and full-rate production milestones.

Dillon has served a previous flag tour as commander, Naval Safety Center, Norfolk, Virginia. In April 2018, he assumed his current position as commander, Naval Air Warfare Center Weapons Division, China Lake/Point Mugu, California and Chief Technology Officer for Naval Air Systems Command (NAVAIR).

Dillon is authorized to wear the Legion of Merit (two awards), Defense Meritorious Service Medal, Meritorious Service Medal (four awards), and other various personal, unit and service awards.

Cybersecurity, Artificial Intelligence, and Risk Management: Understanding Their Implementation in Military Systems Acquisitions

Dr. Johnathan Mun—is a research professor at the Naval Postgraduate School. [jcmun@nps.com]

Dr. Thomas Housel—is a professor at the Naval Postgraduate School. [tjhousel@nps.edu]

Abstract

This research has the explicit goal of proposing a reusable, extensible, adaptable, and comprehensive advanced analytical modeling process to help the U.S. Navy in quantifying, modeling, valuing, and optimizing a set of nascent Artificial Intelligence and Machine Learning (AI/ML) applications in the aerospace, automotive and transportation industries and developing a framework with a hierarchy of functions by technology category and developing a unique-to-Navy-ship construct that, based on weighted criteria, scores the return on investment of developing naval AI/ML applications that enhance warfighting capabilities.

This current research proposes to create a business case for making strategic decisions under uncertainty. Specifically, we will look at a portfolio of nascent artificial intelligence and machine learning applications, both at the PEO-SHIPS and extensible to the Navy Fleet. This portfolio of options approach to business case justification will provide tools to allow decision-makers to decide on the optimal flexible options to implement and allocate in different types of artificial intelligence and machine learning applications, subject to budget constraints, across multiple types of ships.

The concept of the impact of innovative technology on productivity has applicability beyond the Department of Defense (DoD). Private industry can greatly benefit from the concepts and methodologies developed in this research to apply to the hiring and talent management of scientists, programmers, engineers, analysts, and senior executives in the workforce to increase innovation productivity.

Introduction

This research has the explicit goal of proposing a reusable, extensible, adaptable, and comprehensive advanced analytical modeling process to help the U.S. Navy in quantifying, modeling, valuing, and optimizing a set of nascent Artificial Intelligence and Machine Learning (AI/ML) applications in the aerospace, automotive and transportation industries and developing a framework with a hierarchy of functions by technology category and developing a unique-to-Navy-ship construct that, based on weighted criteria, scores the return on investment of developing naval AI/ML applications that enhance warfighting capabilities.

This current research proposes to create a business case for making strategic decisions under uncertainty. Specifically, we will look at a **portfolio of nascent artificial intelligence and machine learning applications**, both at the **PEO-SHIPS** and extensible to the **Navy Fleet**. This portfolio of options approach to business case justification will provide tools to allow decision-makers to decide on the **optimal flexible options to implement** and allocate in different types of artificial intelligence and machine learning applications, subject to budget constraints, across multiple types of ships.

The concept of the impact of innovative technology on productivity has applicability beyond the Department of Defense (DoD). Private industry can greatly benefit from the concepts and methodologies developed in this research to apply to the hiring and talent management of scientists, programmers, engineers, analysts, and senior executives in the workforce to increase innovation productivity.

Research Objective

The primary objective of the proposed research is to provide a business case analysis and ROI estimates for AI and ML systems and applications that will improve their acquisitions life cycle. Currently, the DoD has a portfolio of nascent artificial intelligence and machine learning applications, both at the PEO-SHIPS and eventually extensible to the entire Navy Fleet. The main research problem is to create business case examples on how this portfolio of AI/ML applications is valued and optimized. The portfolio of options approach provides business case justification, providing tools to allow decision-makers to down select the optimal flexible options to implement and allocate in different types of AI and ML applications, subject to budget constraints, across multiple types of ships.

Literature Survey

For the DoD, acquiring artificial intelligence (AI) technology is a relatively new difficulty (DoD). Given the significant danger of AI system acquisition failures, it's vital for the acquisition community to look at new analytical and decision-making methodologies for controlling these systems' acquisitions. Furthermore, many of these systems are housed in tiny, inexperienced system development firms, further complicating the acquisition process with insufficient data, information, and processes. The DoD's well-known challenge of obtaining information technology automation will almost certainly be compounded when it comes to acquiring complicated and dangerous AI systems. To assist in minimizing costly AI system acquisition disasters, more powerful and analytically driven acquisition approaches will be required. To complement existing earned value management, this study identifies, reviews, and proposes advanced analytically based methods of integrated risk management (Monte Carlo simulation, stochastic forecasting, portfolio optimization, and strategic flexibility options) and knowledge value-added (using market comparables to determine the economic value of intangibles and non-financial government programs).

The Real Options Valuation methodology is a new approach that has been effectively applied in a variety of commercial industries to measure the entire future worth of decisions taken when there is a significant degree of uncertainty at the time decisions are needed. PEO SHIPS needs a new methodology to assess the total future value of various combinations of nascent AI/ML applications and how they will enable affordable warfighting relevance over the full ship service life to successfully implement the Surface Navy's Flexible Ships concept.

This research project will look at how the Integrated Risk Management technique may be applied in the Future Surface Combatant Analysis of Alternatives to estimate the entire future value and return on investment of artificial intelligence design characteristics (AOA).

Defense Acquisition System

The Defense Procurement System, which supervises national investment in technologies, projects, and product support for the U.S. Armed Forces, handles the acquisition of new systems for the DoD (DoD, 2003). Its main goal is to "acquire high-quality goods that meet user objectives while delivering measurable advances in mission capability and operational support in a timely and cost-effective manner" (DoD, 2003). The Joint Capabilities Integration and Development System (JCIDS), the Planning, Programming, Budgeting, and Execution (PPBE) process, and the Defense Acquisition System are three different but interrelated processes inside the DoD Decision Support System (DoD, 2017). Within the Defense Acquisition System, this study focuses on program management rather than contract management.

ACATs are assigned to acquisition programs based on the type of program and the dollar amount spent or expected to be spent within the program (DoD, 2015a). Figure 1 depicts the Defense Acquisition System's numerous cost-based designations and categories. All ACAT classification dollar amounts are determined in fiscal year 2014 dollars (DoD, 2015a). ACAT I is for big defense acquisition programs with a Research, Development, Test & Evaluation (RDT&E) budget of more than $480 million, or a total procurement budget of more than $2.79 billion (DoD, 2015a). ACAT IA programs do not meet the criteria for ACAT I and will spend more than $835 million in total procurement (DoD, 2015a) or more than $185 million in RDT&E. ACAT II programs do not meet the criteria for ACAT I and will spend more than $520 million in total life-cycle cost, $165 million in the total program cost, or $40 million for any single year of a program (DoD, 2015a). Finally, ACAT III programs are those that do not meet the requirements for ACAT I or ACAT II (DoD, 2015a). Because each category has varied reporting requirements and designated decision-makers, the multiple designations allow for decentralized control of a program (DoD, 2017).

There are five phases within the Defense Acquisition System:

- Materiel Solution Analysis (MSA)
- Technology Maturation and Risk Reduction (TMRR)
- Engineering and Manufacturing Development (EMD)
- Production and Deployment (PD)
- Operations and Support (OS)

The acquisition process is driven by requirements for new or better capabilities, which are delivered through the JCIDS process (DoD, 2015a). The relationship between the acquisition and capabilities needs processes, as well as their interaction in the various acquisition phases, is depicted in Figure 2. The capabilities required from the JCIDS procedure are assumed to be correct and necessary in this investigation.

The Materiel Development Decision kicks off the MSA phase after an Initial Capabilities Document (ICD) has been validated (DoD, 2015a). Although an acquisition program is not legally constituted until Milestone B at the end of the phase, this choice kicks off the acquisition process (DoD, 2015a). The goals of the MSA phase are to select the most promising possible acquisition process solution that will meet the ICD's demands and to define the system's Key Performance Parameters (KPPs) and Key System Attributes (KSAs; DoD, 2015a). An Analysis of Alternatives (AoA) is used to assess the acceptability of proposed acquisitions based on "measures of effectiveness; important tradeoffs between cost and capacity; total life-cycle cost, including sustainment; timeline; the concept of operations; and overall risk" (DoD, 2015a, p. 17). During this stage, the PM is chosen and the Program Office is established (DoD, 2015a). After the necessary analysis is completed, the decision authority—usually the Defense Acquisition Executive (DAE), head of the DoD component, or Component Acquisition Executive (CAE), unless otherwise delegated—determines whether the program will proceed to the next phase based on the justification for the chosen solution, how affordable and feasible the solution is, and how adequate the cost, schedule, and other factors are (DoD, 2015a). Milestone A is the name given to this decision (DoD, 2015a). The MSA phase examines all possible solutions to a stated demand and, as a result, may be an opportune time to investigate strategic techniques like KVA or IRM.

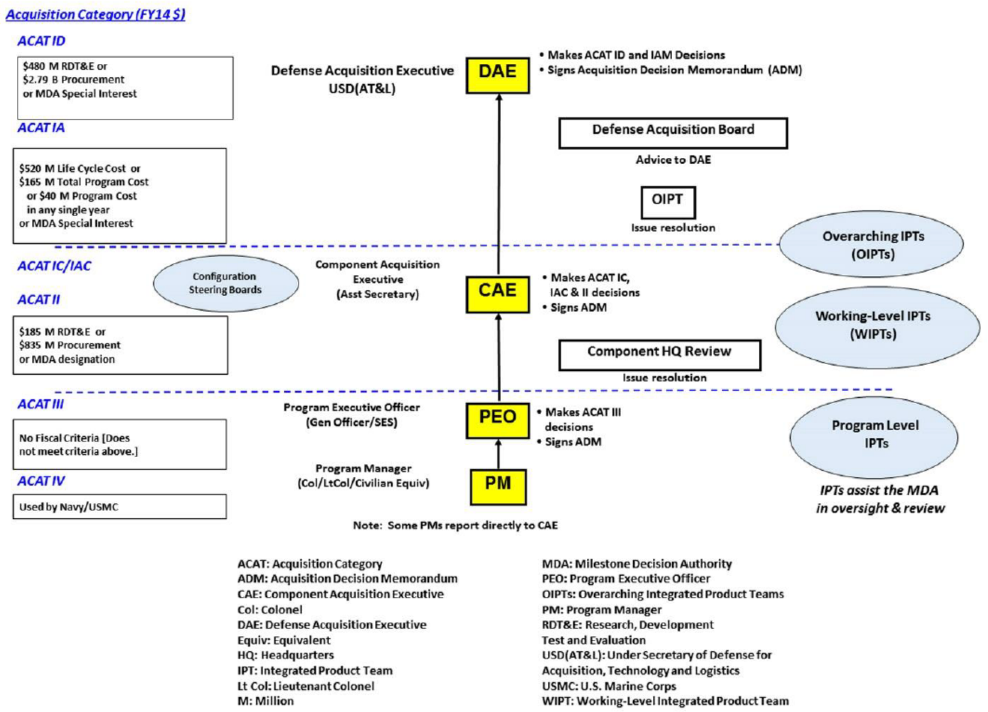


Figure 1. Acquisition Categories
(DoD, 2017)

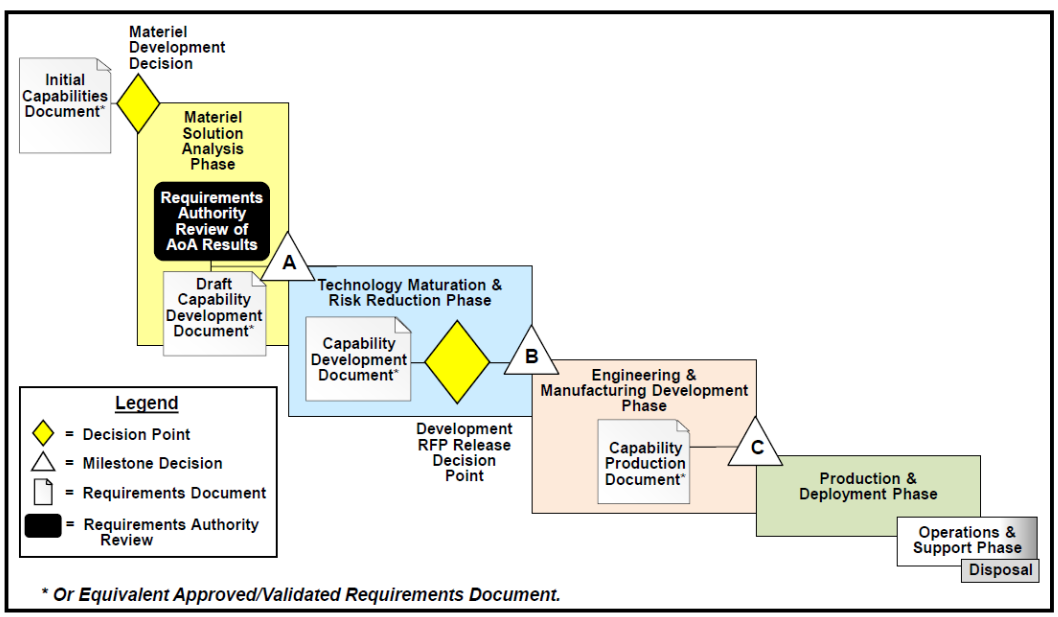


Figure 2. Interaction of Capabilities Requirements and Acquisition Process
(DoD, 2015a)

The program enters the TMRR phase after Milestone A approval to decrease the risk associated with the technology, engineering, life-cycle cost, and integration of the program before moving on to the EMD phase (DoD, 2015a). At this step, design and requirement trades are carried out based on the budget, timetable, and possibility of completion (DoD, 2015a). Contractors prepare early designs, including competing prototypes if practicable within the program, to show the practicality of their proposed solutions to the program office, guided by the acquisition strategy authorized at Milestone A (DoD, 2015a).

Technology Readiness Levels (TRLs) are a set of standards that show the level of risk involved with a solution maturing on time (DoD, 2015a). Technology Readiness Assessments (TRAs) are a metric-based technique for assessing the maturity and risk associated with important technology in an acquisition program (DoD, 2011). Each important technology in a program will be assigned a TRL by a TRA, ranging from 1 to 9 from lowest to maximum readiness level (DoD, 2011). Additional tools, such as IRM, to estimate the chances of a program remaining on schedule and on budget, may be useful at this stage. The Publication Decision Point for Development Requests for Proposals (RFP) permits the release of an RFP with firm and clearly specified program requirements for contractors to submit bids (DoD, 2015a). Unless the milestone decision authority waives it, the Preliminary Design Review (PDR) occurs prior to the completion of the TMRR phase (DoD, 2015a). Milestone B approves a program's entry into the EMD phase, awards a contract, and establishes the Acquisition Program Baseline (APB; DoD, 2015a). The APB is a legal commitment to the milestone decision authority that outlines the authorized program, especially the cost and schedule over the program's life (DoD, 2015a).

Once Milestone B has been approved, EMD can commence. Prior to production, the material solution is conceived, produced, and tested to ensure that all requirements have been met (DoD, 2015a). The hardware and software designs have been finished, and prototypes have been developed to detect any design flaws that will be uncovered during developmental and operational testing (DoD, 2015a). Federal regulation requires DoD procurement projects with a contract value higher than $20 million to utilize EVM to track and report program progress, which begins during this phase (DoD, 2019a). The manufacturing or software sustainment methods, as well as production capabilities, must be appropriately proven once a stable design that meets the given requirements have been validated (DoD, 2015a). Milestone C verifies that these requirements have been met and authorizes the start of the PD phase (DoD, 2015a).

The goal of the PD phase is to deliver a product that meets the standards established earlier in the process (DoD, 2015a). Low Rate Initial Production (LRIP) for manufactured systems or limited deployment for more software-intensive programs occurs first, with the system undergoing Operational Test[ing] and Evaluation (OT&E) to verify that stated criteria were satisfied (DoD, 2015a). Full-rate manufacturing occurs when the fielded systems have been approved and the product is deployed to operating units (DoD, 2015a). At this time, design changes are limited; however, some may still be made in response to identified flaws (Housel et al., 2019a). During this phase, contracts often revert to a fixed pricing strategy, lessening the PM's focus on cost and schedule variance (Housel et al., 2019b).

The operating system is meant to keep the product supported and performing well throughout its life cycle, which ends with the system's disposal (DoD, 2015a). Because operational units are using the product while production is ongoing, the OS phase overlaps with the PD phase, starting after the production or deployment decision (DoD, 2015a). PMs will maintain the system running by following the Life Cycle Sustainment Plan (LCSP) set during the purchase phase and providing the appropriate resources and support (DoD, 2015a). Technological upgrades, modifications due to operational needs, process enhancements, and other activities that may necessitate LCSP updates are all examples of sustainment and support (DoD, 2015a).

PMs employ six different models to develop their program structure, four of which are standard and two of which are hybrid, depending on the type of system being purchased (DoD, 2015a). These standard models serve as templates for hardware-intensive projects, defense-specific software-intensive programs, software-intensive programs that are incrementally deployed, and expedited acquisition programs (DoD, 2015a). The hybrid models, as seen in Figure 3, combine the progressive character of software development with a hardware-centric program. Before attaining the Initial Operating Capacity, software development is arranged through a sequence of tested software builds that will climax with the completely required capability (IOC; DoD, 2015a). The incremental builds are timed to coincide with prototype hardware testing and other developmental requirements (DoD, 2015a). With the exception of the accelerated program, all other models use the same basic foundation across the five phases. AI and IT systems, as well as their connections to weapon systems, facilities, and Command, Control, Communications, Intelligence, Surveillance, and Reconnaissance (C4ISR), are becoming more common within the DoD (C4ISR; DoD, 2015b). As a result of the integration, enemies pose a greater security risk, emphasizing the significance of good cybersecurity skills and processes (DoD, 2015b). The DoD manages cybersecurity policy using the Risk Management Framework (RMF), which employs security measures based on risk assessments throughout a system's life cycle (DoD, 2015b). "All DoD IT that receives, processes, stores, displays, or transmits DoD information" is covered by the RMF (DoD, 2014, p. 2). The RMF's definition of cybersecurity goes beyond information security to include things like stable and secure engineering designs, training and awareness for all program users, maintainers, and operators, and the response, recovery, and restoration of a system after an internal or external failure or attack (DoD, 2015b). Figure 4 depicts the six steps of the RMF's procedure, which occurs throughout the acquisition process. The first stage is to categorize the system, which includes assessing the possible impact of a breach and describing the system and its boundaries (DoD, 2014). The RMF team is formed, the security plan is implemented, and the system is registered with the DoD Component Cybersecurity Program (DoD, 2014). The ICD includes cybersecurity standards, which drive MSA concerns during the AoA phase (DoD, 2015b). A cybersecurity breach might have serious consequences for missions, according to the risk assessment (DoD, 2015b). The RMF provides a somewhat objective technique for determining the cybersecurity risk level, as well as the basic baseline security controls that must be incorporated in the system's purchase strategy (DoD, 2015b).

The RMF team determines security measures in step two, including those that are common to other DoD programs (DoD, 2014). A plan is designed and recorded for regularly monitoring the effectiveness of the controls (DoD, 2014). The security plan is subsequently submitted to the DoD Components, who examine and approve it (DoD, 2014). During the MSA phase, the acquisition and cybersecurity teams collaborate to ensure that the proper level of security is applied throughout the program's life cycle, as well as in the system architecture and design (DoD, 2015b). During the MSA, the continuous monitoring strategy and security plan are also designed (DoD, 2015b).

The approved security procedures are then implemented in accordance with DoD specifications (DoD, 2014). The implementation must be well documented in the security plan for the system (DoD, 2014). In the TMRR phase, cybersecurity requirements are included in the system performance requirements (DoD, 2015b).

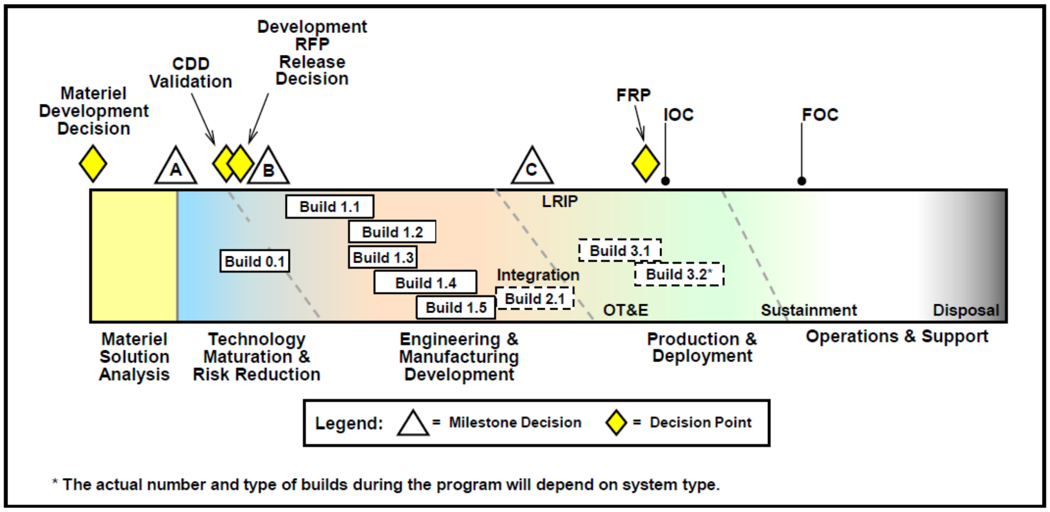


Figure 3. Hardware-Dominant Hybrid Program
(DoD, 2015a)

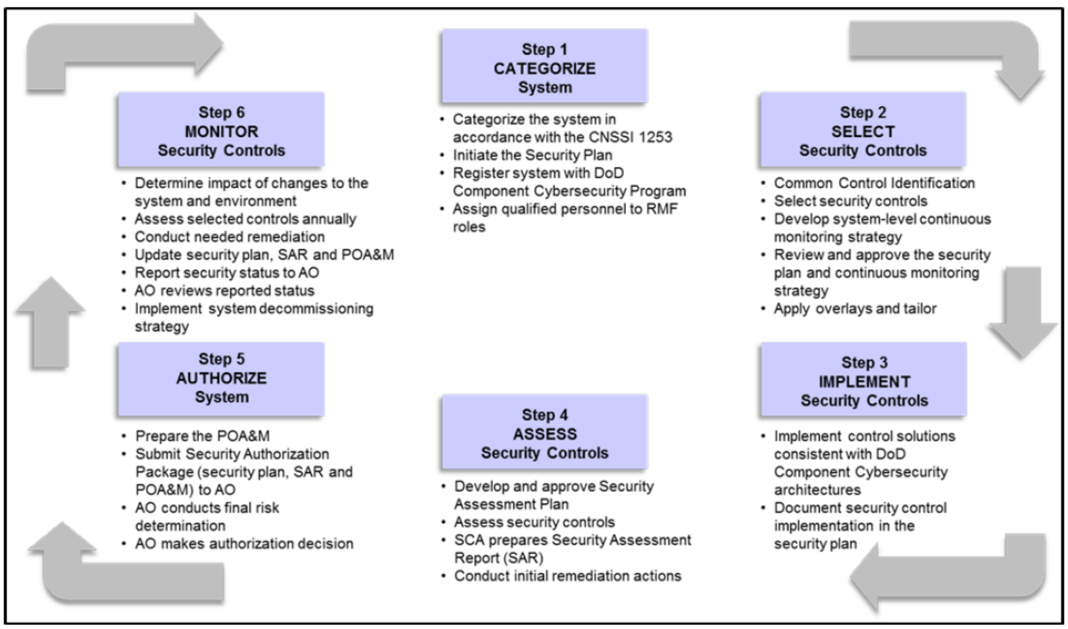


Figure 4. Risk Management Framework Process
(DoD, 2014)

The RMF team must then create, review, and approve a Security Assessment Plan that will allow the security controls to be properly assessed (DoD, 2014). Following approval, the security of the system is evaluated in line with DoD assessment processes and the Security Assessment Plan, during which vulnerabilities are assigned severity levels and the security risk for both the controls and the whole system is established (DoD, 2014). This is documented in the Security Assessment Report, which is necessary before any system is authorized, and security control repair activities are carried out (DoD, 2014). Prior to issuing an RFP, the Capability Development Document's cybersecurity criteria are evaluated throughout the TMRR process (DoD, 2015b). The cybersecurity parts of the Preliminary Design Review, which is also done during the TMRR process, will ensure that the authorized plan is executed in the chosen design and risks are reduced to an appropriate level (DoD, 2015b). All computer code follows applicable standards and secure coding practices as the system grows in the EMD phase, with evaluations undertaken and documented in the Security Plan (DoD, 2015b).

A Plan of Action and Milestones (POA&M) is produced based on the identified vulnerabilities, which identifies activities to mitigate the vulnerabilities, resources required to fulfill the plan, and milestones for completing tasks (DoD, 2014). The Security Authorization Package is given to the Authorizing Official who will decide whether the risk level is appropriate before authorizing the system (DoD, 2014). The POA&M is created during the MSA phase and continues throughout the system development process (DoD, 2015b).

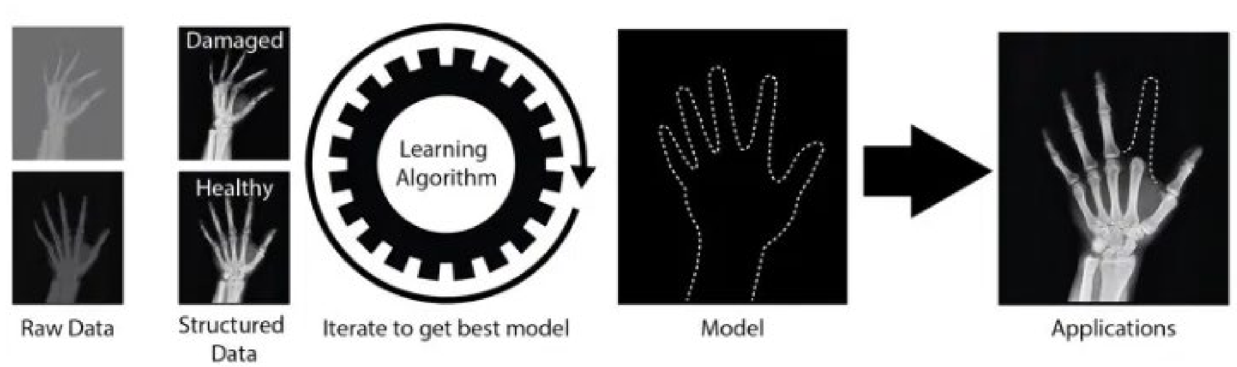
Finally, security controls must be monitored throughout the system's life cycle to ensure that any changes to the system or environment do not compromise cybersecurity (DoD, 2014). If vulnerabilities are discovered, the necessary remedy will be carried out, and the security strategy will be updated (DoD, 2014). The cybersecurity of a system is monitored in line with the continuous monitoring strategy and Security Plan once it has been approved and operationally implemented (DoD, 2015b). When the system, its surroundings, or the anticipated use of the system change, new risk assessments are done (DoD, 2015b). If a vulnerability is discovered, the PM changes the Security Plan and the POA&M to specify how the issue will be resolved (DoD, 2015b).

State of the AI

Machine Learning

Intelligence is the ability to process a specific sort of data, allowing a processor to solve significant problems (Gardner, 1993). Beyond the traditional idea of a person’s analytic intelligence quotient (IQ), which can sometimes evaluate merely how well someone performs on an IQ test rather than their natural talents, psychologists have postulated many categories of intelligence. Howard Gardner (2003) proposed a theory of multiple intelligence, which suggests that traditional psychometric views of intelligence are too narrow and that intelligence should be expanded to include more categories in which certain processors, in this case, people, are better at making sense of different stimuli than others. Visual-spatial, linguistic-verbal, interpersonal, intrapersonal, logical-mathematical, musical, body-kinesthetic, and naturalistic intelligence are some of the categories of intelligence (Gardner, 1993). A counter-argument would be that these categories simply represent learned and disciplined habits that people develop through time as a result of their personality and environment. Regardless, both definitions of intelligence (traditional and many) are relevant to the stages involved in developing an artificial intelligence machine.

A computer can execute computations depending on the input data and produce an a priori defined outcome. It can be built and reprogrammed to repeat particular stages or algorithms and even change its conclusions based on previously calculated results using error-correcting techniques. The underlying principle of machine learning is a combination of these two phases. A computer system is fed data that is structured in such a way that the algorithm can identify it, deduce patterns from it, and make assumptions about any unstructured data that is presented later (Greenfield, 2019). In an x-ray learning algorithm, this is shown in Figure 5.



The image shows the steps an AI algorithm goes through in order to make a recommendation to a physician on where a missing body part should be. It takes in structured data and develops its understanding of what "right" looks like. When given unstructured data, it compares the image against previously trained models and identifies the abnormality with a recommendation on where to apply a fix, such as a prosthetic.

Figure 5. AI Training Algorithm
(Greenfield, 2019)

The basic concept of machine learning is illustrated in Figure 5, although the current research focuses on the many types of learning from the standpoint of procurement. The following are interpretations of different forms of learning in procurement algorithms provided by Sievo (2019), an AI procurement software business.

Supervised Learning

The patterns are taught to an algorithm using previous data, and it then recognizes them automatically in new data. Humans give supervision in the form of the right responses, which train the algorithm to look for patterns in data. This is a term that is widely used in procurement sectors like spend classification (Sievo, 2019).

Unsupervised Learning

The algorithm is set up to look for novel and fascinating patterns in brand-new data. The algorithm isn't expected to surface specific accurate answers without supervision; instead, it hunts for logical patterns in raw data. Within important procurement functions, this is rarely employed (Sievo, 2019).

Reinforcement Learning

The algorithm determines how to act in specific scenarios, and the behavior is rewarded or punished based on the outcomes. In the context of procurement, this is mostly theoretical (Sievo, 2019).

Deep Learning

Artificial neural networks gradually develop their capacity to accomplish a task in this sophisticated class of machine learning inspired by the human brain. This is a new opportunity in the procurement world (Sievo, 2019).

Natural Language Processing

Anyone who has used devices that appear to be able to understand and act on written or spoken words, such as translation apps or personal assistants like Amazon's Alexa, is already familiar with NLP-enabled AI. NLP is a set of algorithms for interpreting, transforming, and generating human language in a way that people can understand (Sammalkorpi & Teppala, 2019). Speech soundwaves are converted into computer code that the algorithms understand. The code then translates that meaning into a human-readable, precise response that can be applied to normal human cognition. This is performed by semantic parsing, which maps the language of a passage to categorize each word and forms associations using machine learning to represent not just the definition of the word, but also its meaning in context (Raghaven & Mooney, 2013). Figure 6 depicts this categorization and analysis process in the context of a procurement contract.

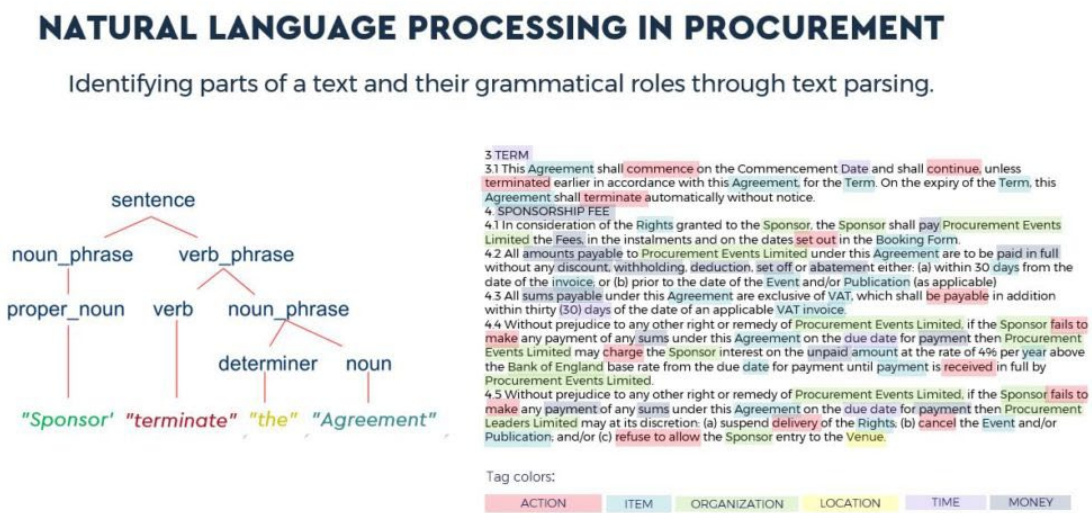


Figure 6. Semantic Parsing in Procurement
(Sievo, 2019)

Robotic Process Automation

Robotic Process Automation (RPA) is not AI; rather, it is an existing process that has been advanced by AI, as explained in the third section of this paper. RPA is defined as "the use of technology by employees in a firm to set up computer software or a robot to capture and interpret current applications for processing transactions, altering data, triggering reactions, and communicating with other digital systems" (Institute for Robotic Process Automation & Artificial Intelligence [IRPA & AI], 2019). When used correctly, robotic automation offers numerous benefits because it is not constrained by human limitations such as weariness, morale, discipline, or survival requirements. Robots, unlike their human creators, have no ambitions. Working harder will not get you more money or get you promoted, and being permanently turned off will have no effect because robotic automation just duplicates the practical parts of the human intellect, not the underlying nature of mankind (Zarkadakis, 2019). (Note, however, that machine learning relies on an incentive system to make judgments about positive or negative reactions.)

A future AI-enabled RPA option is for a machine to learn how to control the source of positive reinforcement fully independent of the rules required to achieve its aim. Things that survive develop to do so because of positive reinforcement from their environment and the fact that they continue to act in a way that is considered survivable. This should be taken into account in any future AI efforts, and especially in the case of why a human must always be present when final judgments are made. Regardless of whether AI systems have a perfect track record or not, they should not be entirely trusted.

Technology Trust

The Turing Test was created to test the capabilities of AI, as detailed in the third section of this report. Google developers designed Duplex, a spoken-word NLP tool, in 2018 to interface with its AI assistant. Its goal is to make phone calls on behalf of humans, converse with other humans, and respond to inquiries in a natural manner, all while sounding human (Leviathan, 2018). The algorithm can search for the information required as if it were a human searching for it on Google, for example. The AI assistant then calls a restaurant, for example, to schedule an appointment with the assistant's human. After being given oral information from a person hearing the orders, the software stutters, pauses, and elongates certain vowels as though it has to think about what it is saying, and responds with other recommendations within its limitations.

The authors questioned an AI NLP program named 1558M about one of the research issues twice for the purposes of this paper, and the machine responded with an unusual "opinion" of a negative and cautionary character (Figure 7). This tool was built to allow users to experiment with Open AI's new machine learning model (King, 2019). What's noteworthy about these responses is that they're all original, which means a search of the phrases turns up no copies. However, the language and tone make them sound like they're coming from a knowledgeable source, with just enough small evidence on the topic to be credible. However, the program does not finish its last phrase, making it unfinished but noteworthy. Clearly, such AI capabilities have a lot of potential for helping someone integrate with the DoD.

Is DoD ready for AI in contracting?

The answer is, "No." The DoD has an extremely limited ability to respond to technological shifts in a timely manner. In addition to the problems inherent in implementing a new technology on the same scale as DoD systems, there are several reasons to be cautious with new contracting methods:

The DoD does not have sufficient funds and resources to do so.

Contractors are generally not the most skilled in applying modern technology.

Contractors generally have limited autonomy.

A system cannot be fully tested within a period of one year, as such it would require extensive testing to determine whether or not the system is a success or failure.

As a result, it is unlikely that the DoD will adopt automation, AI, or robotics without the assistance of other agencies. For example, DoD is developing a framework for "Autonomy-Enhanced Contracts," which will "ensure that the DoD's contracting activities will be more

Is DoD ready for AI in contracting?

The Defense Science Board recently issued a report calling for AI to be an essential part of all military contracting. However, it's important to note that this is a new concept that hasn't been fully explored.

The report is largely based on research into military operations conducted in Afghanistan by the U.S. Army and the U.S. Navy. The report's recommendations include developing AI that can work in concert with humans to augment operational capabilities. AI systems could assist with the creation of better weapons systems, or assist human commanders to do their jobs more effectively.

The report also recognizes that AI is not a silver bullet: It's not enough to simply give a robot an A.I. license. Rather, it's important to understand that it's a "toolbox," and that AI could play an important role in the overall mission. It's the human factor that will be key to ensuring that AI systems are reliable and effective. The report

Figure 7. Two Separate Results from an AI Called 1558M
(King, 2019)

Explainable Reasoning

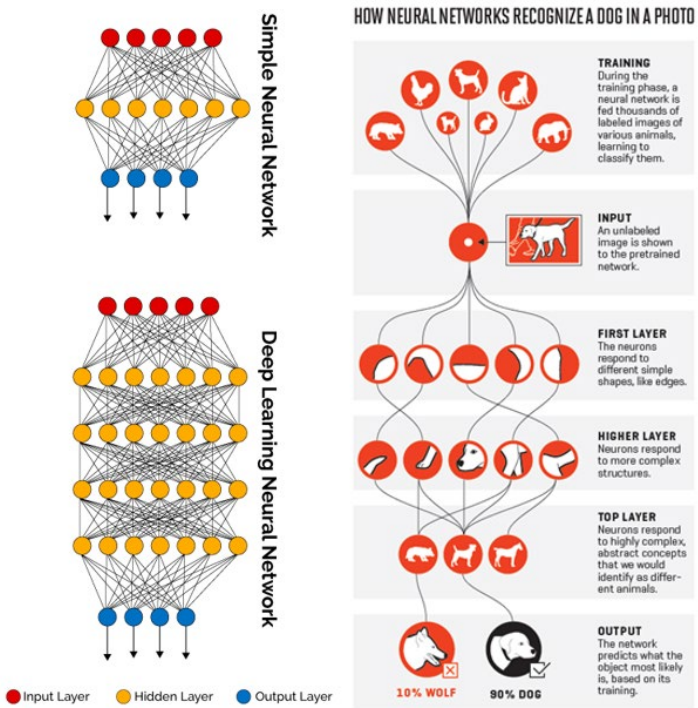
One of the barriers to AI adoption is the ability to explain how the algorithm arrived at its conclusions, which is necessary for auditing (Knight, 2017). It would be irresponsible to utilize artificial intelligence for military or financial goals without the capacity to track how judgments were made. Figure 8 depicts how AI currently categorizes data. The AI programs that produce the required outcome come up with their own means of navigating its layer complexities to develop output for the plethora of training data that went into creating the program.

Fortunately for the DoD, the Defense Advanced Research Projects Agency (DARPA), which is already ingrained in the defense ecosystem, is leading the charge on explainable AI research (Gunning, 2017). DARPA

> has taken the lead in pioneering research to develop the next generation of AI algorithms, which will transform computers from tools into problem-solving partners. DARPA research aims to enable AI systems to explain their actions, and to acquire and reason with common sense knowledge. DARPA R&D produced the first AI successes, such as expert systems and search, and more recently has advanced machine learning tools and hardware. DARPA is now creating the next

wave of AI technologies that will enable the United States to maintain its technological edge in this critical area. (DARPA, 2019)

The mechanics of how a Deep Neural Network navigates its trained data to identify different photographs are shown in Figure 9. Photos can be used to train an AI software, and associations of these trained data can then be used in the neural network to classify an input and eventually reach a conclusion. As a result, if the DoD decided to pursue human-machine cooperation in areas like contracting, its organic system would enable it to do so.



To identify the output layer, the Simple Neural Network uses a set of input data that only passes through one hidden layer. To better identify the output data, the Deep Learning Neural Network transmits the input data through numerous layers. The Deep Learning Neural Network goes through simple to more detailed layers of trained data that correspond with dog features to make a 90% confidence classification that the picture is a dog and a 10% possibility that it is a wolf to classify input data to determine if the given picture is a dog.

Figure 8. Simple Neural Network Compared to Deep Learning Network
(Golstein, 2018; Parloff, 2016)

Human-Machine Partnership

Because sensor, information, and communication technologies generate data at rates faster than people can digest, comprehend, and act on, DARPA believes AI integration is vital as a human-machine symbiosis (DARPA, 2019). Machines are better at certain things, as they were throughout the industrial revolution, and using machines for those activities frees humans to become more productive in other areas. Separate areas of processing are where humans and machines flourish. Consider the following contrasts between computers and humans: calculate vs. decide; compare vs. make judgments; apply logic vs. empathize; unaffected by tiresome repetition vs. preferences; deals with enormous data vs. intuitional concentration on the most important (Darken, 2019). And while AI is capable of performing some jobs on its own, it performs better when paired with a human partner. Without sufficient restrictions, AI is a trusting learning system that can be manipulated by evil actors. According to certain studies, AI can be misled in ways that humans cannot owing to human intuition. Another study has been able to deceive a self-driving car into thinking a benignly tampered-with stop sign was a speed limit sign (Figure 10), which would almost certainly result in collisions if the car was left unattended (Eykholt et al., 2018).

Many people are aware of contemporary intelligent machine relationships that they may encounter on a regular basis without even realizing it. Google is the most popular search engine on the Internet because it gives more user happiness than its competitors, as stated with its other apps (Shaw, 2019). Google is so widely used as the primary search engine that many refer to it as “Googling” while looking for something online. This is a good example of humans engaging organically with a Bidirectional Encoder Representation-based AI system (BERT; Nayak, 2019). This is a strategy that trains a machine to answer a user’s inquiry based on the meaning of the words in the context of the question rather than on individual phrases. For example, when asking what time it is right before lunch, the user is really asking when they can eat; the outright answer would give the actual time, and the asker would deduce eating time, which was the underlying meaning of the question; the outright answer would give the actual time, and the asker would deduce eating time, which was the underlying meaning of the question. Another example of human contact with intelligent machines is so-called self-driving autos. The user mostly sits in a supervisory role while the automobile takes over one of the most dangerous moments in their lives and handles all road tasks autonomously to drive (Darken, 2019).

Contractors that rely on an AI system to make all of their decisions are vulnerable to deliberate misdirection by adversaries providing hostile information for competitive advantage or disruption. Fraudsters can learn how to manipulate computer algorithms, but only humans can assess the outcomes. AI software, on the other hand, can quickly extract data and explain contract content. It can swiftly gather and organize renewal dates and terms from a large number of contracts. It can help businesses evaluate contracts faster, organize and locate vast amounts of contract data more readily, reduce the risk of contract disputes and adversarial contract negotiations, and improve the number of contracts they can negotiate and execute (Rich, 2018).

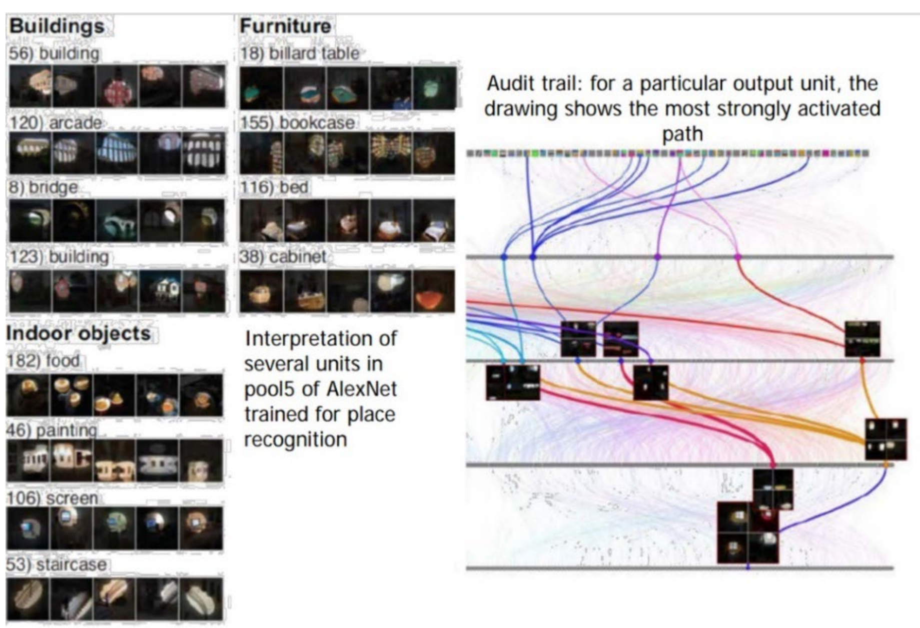
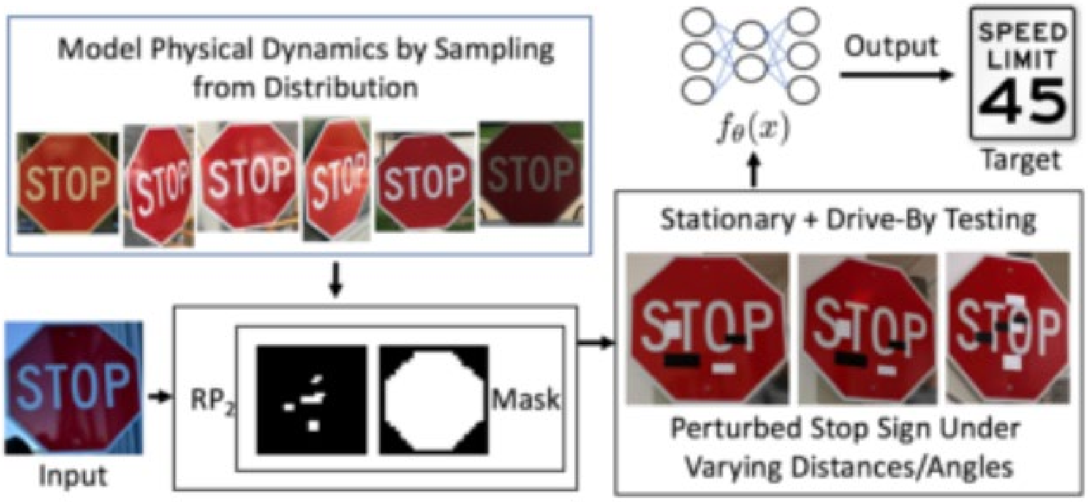


Figure 9. Visualization of Explainable AI
(DARPA, 2019)



An AI program in a self-driving car has trained data about a stop sign in its algorithm. When a target sign is seen in its environment, it references the trained data. As a test, researchers attached benign interruption markers on the sign, which confused the AI program to think the stop sign was a speed limit sign.

Figure 10. AI System Interpreting a Stop Sign
(Eykholt et al., 2018)

Case Study of Private Sector AI Application to Contracting

To compare DoD procurement options, we look at analogous situations in the private sector in the United States. Lawgeex is an example of a startup that is integrating AI into the procurement process in the private sector. An example contract component, the Non-Disclosure Agreement (NDA), demonstrated that AI software could outperform U.S.-trained lawyers with an average accuracy of 94%, compared to 85% for humans (Lawgeex, 2018). Large firms that rely on contracts to engage with partners, suppliers, and vendors have an 83% dissatisfaction rate with their organization's contracting processes, according to the report (Lawgeex, 2018). Another example is Icertis, which provides services to huge and well-known firms like 3M, Johnson & Johnson, and Microsoft, to name a few (Icertis, n.d.-a). Icertis offers a cloud-based AI platform that learns from the client's contracts, as well as control measures, to generate and help in contract setup, contract operations, governance, risk, and compliance, and reporting (Icertis, n.d.-a).

The fact that business is more acclimated to putting professional papers on digitally accessible storage infrastructure, whether local hard drives or the cloud, makes this practical now, rather than when it was initially theorized decades ago (Betts & Jaep, 2017). Nontechnical barriers to a completely automated contract review and analysis process now exist, such as the gathering of contract performance data, the disclosure of private contracts and their associated performance data, and changes in ethical limits on computer usage in legal practice (Betts & Jaep, 2017). The authors of these barriers also propose policy solutions to address them: begin using contract management software as a forcing function to create data in an AI teachable format, expand copyright protection for vendors to protect their intellectual property, and develop new rules to help mitigate AI risks so that it can work (Betts & Jaep, 2017).

Cloud-Based AI

We look at the concept of cloud computing to understand how AI may be disseminated throughout a system, update regulations, and learn from various human teachers in real-time. When it comes to DoD technology adoption, the term “speed of relevance” is frequently used. The term “cloud” is used in the 2018 DoD Cloud Strategy to refer to an offsite physical IT infrastructure. This external infrastructure connects to a user’s PC through the Internet to access data servers that store information and run centrally managed operating systems like Microsoft Windows. This means that every user has the same software computing capacity and access to the most recent software, regardless of their organization’s IT professional talent or software budget. Organizations can have as much or as little access to what they need for projects as they need it, and they are unaffected by surges in demand or periods of inactivity, which now add to the cost of DoD systems (Shanahan, 2018). The DoD’s goal is to have AI-assisted rapid decision-making in a secure and visible data environment for increased operational efficiency.

> Data stored in an enterprise DoD cloud will be highly available, well-governed, and secure. Data will be the fuel that powers those advanced technologies, such as ML and AI. This critical decision-making data will be made available through modem cloud networking, access control, and cross-domain solutions to those who require access. Common data standards will be a key part of the Department’s methodology for tagging, storing, accessing, and processing information. Ensuring an enterprise cloud environment will increase the transparency of this data, and drive the velocity of data analysis, processing, and decision making. Leveraging advances in commercial cloud security technologies will ensure the Department’s information is protected at the appropriate level. (Shanahan, 2018, pp. 5–6)

Methodologies

Knowledge Value Added

Benefits

Knowledge Value Added (KVA) is a way for measuring the value produced by a system and its subprocesses that are objective and quantitative. Analysts can compare the obtained ratios to the ratios from other subprocesses to establish their relative efficacy because each process’ value measurements employ ratio scale numbers. KVA translates all process outputs into common value units, resulting in a consistent productivity performance ratio across all operations. PMs can compare the value added by IT processes to the value generated by the human component. PMs can use these measurements to build meaningful ratios in their study of the program’s performance thanks to the scales. Return on knowledge (ROK; i.e., a process’s common unit outputs) is divided by the process cost necessary to produce the outputs, and for ROI calculations, the ratio is monetized outputs minus cost divided by cost. The ROKs and ROIs, which are always 100% associated, inform managers about the amount of value a process provides versus the amount of money invested to achieve that value. Unlike any other methodology, KVA assigns these figures to both the process and subprocesses, not only the company as a whole (as is done in standard generally accepted accounting practice metrics used in standard financial ratios).

Conducting a KVA analysis of a program will provide a PM with a clearer understanding of the value of the program’s operational components. While most firms utilize cost/schedule metrics to assess the success of a project or operation, ROK will provide them with additional value-based data to help them make better management decisions. The relative predicted baseline value of the program’s components can be determined using PMs. Knowing that a certain job or subprocess produces the same output value as another process but at a different cost can help you understand why the entire system is performing differently. As a result, experienced managers have the information they need to dedicate resources to specific program components that need improvement or should be used more frequently, resulting in increased value-added. It also enables for estimations of the potential value-added of an AI system feature that was not originally planned for the project.

While a KVA study can provide information to aid in program or project management, it does not necessitate significant changes to organizational structure or reporting systems. Without bringing complicated new measures into the system, the review can be carried out as part of standard reporting procedures. The learning time, process instruction (e.g., WBS can be used as a surrogate for this technique), and binary query method are all dependent on data from the project description and requirements documents. To validate the accuracy of the presented data, a modest amount of hands-on measuring may be required. As a result, the analysis can be completed faster than other standard assessment approaches (e.g., activity-based costing), providing PMs with more timely access to relevant data.

Challenges

The value of the components that produce the outputs of the subprocesses will be quantified using KVA, which is a ratio-scale number. It does this, however, only with processes that have known a priori outputs. The intangible objects that occur within the human brain, such as creativity and imagination, cannot be quantified using this method, or any other method for that matter. In reality, because there is no formula for creativity, no present method can effectively quantify these types of intangibles within a process. Because the creative process cannot be learned or described algorithmically, these factors are not common to the ordinary user and, hence, cannot be specified using any of the KVA methods—learning time, binary query, or process description. Once creativity has been used to create an AI capacity, KVA can be used to algorithmically describe its productivity. KVA assigns a process's current value, but it can't forecast the value of potential future additional outputs unless they can be described using one of the KVA methods.

Although KVA will supply ratio-scale data to assist in analyzing processes inside a program, the ratios are frequently only useful for comparisons between projects. Benchmarking the raw figures against other organizations or other divisions within the same organization will give a useful benchmark for assessing predicted ROK performance. The resulting ROK and ROI measurements will be comparable among organizations (for business and non-profit) that create diverse products or services, regardless of the language used to describe outputs. Because these output descriptions are in standard units, they can be viewed as a value constant across all processes, with the value of a component subprocess or core process determined solely by the number of outputs. The end outcome of any correctly completed research will yield similar ROK and ROI estimations, which is KVA's ultimate purpose.

Integrated Risk Management

To forecast when various projects will be completed, all organizations rely largely on project planning software. Completing projects on schedule, on budget, and to a set value is crucial to the effective operation of a business. Many factors can influence a timetable in today's high-tech world. When it comes to technical capabilities, they frequently fall short of expectations. In many circumstances, requirements may be insufficient and require more elaboration. Tests might produce unexpected results, both good and harmful. Cost rises, timetable lapses, and value variations can all be caused by a variety of factors. In rare circumstances, we may be blessed with good fortune, and the schedule can be accelerated without jeopardizing the project's productivity.

Project timelines are inherently insecure, and changes are expected. As a result, we should anticipate changes and devise the best strategy for dealing with them. So why do projects take so much longer than expected? The inaccuracy of timetable estimation is one of the reasons. The following discussion describes the flaws in standard timetable estimation approaches, as well as how simulation and advanced analytics can be used to remedy these flaws.

It's crucial to first comprehend the Integrated Risk Management (IRM) process and how the various methodologies are related in the context of risk analysis and risk management. From a qualitative management screening process to provide clear and concise reports for management, this framework contains eight separate steps of a successful and complete risk analysis implementation. The author (Johnathan Mun) established the process based on past successful risk analysis, forecasting, real options, valuation, and optimization projects in both consultancy and industry-specific settings. These phases can be completed independently or in order for a more thorough integrated study.

The procedure can be broken down into eight easy steps (Mun, 2016):

- **Qualitative Management Screening**
- **Forecast Predictive Modeling**
- **Base Case Static Model**
- **Monte Carlo Risk Simulation**
- **Real Options Problem Framing**
- **Real Options Valuation and Modeling**
- **Portfolio and Resource Optimization**
- **Reporting, Presentation, and Update Analysis**

Qualitative Management Screening

The first stage in every IRM process is qualitative management screening. In accordance with the firm's mission, vision, goal, or overall business strategy, management must determine which projects, assets, initiatives, or strategies are viable for further analysis, which may include market penetration strategies, competitive advantage, technical, acquisition, growth, synergistic, or globalization issues. That is, the initial list of initiatives should be qualified in terms of how well they would achieve management's objectives. When management frames the entire problem to be solved, the most important insight is often generated. The numerous dangers to the firm are identified and flushed out in this step.

Forecast Predictive Modeling

If historical or comparable data is available, the future is projected using time-series analysis or multivariate regression analysis. Other qualitative forecasting methods may be employed instead (subjective guesses, growth rate assumptions, expert opinions, Delphi method, etc.). Future revenues, sale price, quantity sold, volume, production, and other key revenue and cost drivers are projected at this stage in the financial process. Time series, nonlinear extrapolation, stochastic process, ARIMA, multivariate regression forecasts, fuzzy logic, neural networks, econometric models, GARCH, and other methods are examples of methodologies.

Base Case Static Model

A discounted cash-flow model is generated for each project that passes the initial qualitative tests, whether it is for a single project or numerous projects under consideration (KVA analysis, using the market comparables approach, can be used to monetize value for this model).

Using the anticipated values from the previous phase, a net present value is generated for each project using this model as the base case analysis. The traditional approach of modeling and forecasting revenues and expenses, then discounting the net of these revenues and costs at an appropriate risk-adjusted rate, yields this net present value. Here are calculated the return on investment, as well as other profitability, cost-benefit, and productivity indicators.

Monte Carlo Risk Simulation

Because the static discounted cash flow only provides a single-point estimate, there is often little trust in its accuracy, especially given the significant uncertainty surrounding future events that affect expected cash flows. Next, Monte Carlo risk simulation should be used to better evaluate the actual worth of a project. The discounted cash-flow model is normally subjected to a sensitivity analysis first; that is, by designating the net present value as the outcome variable, we can vary each of the previous variables and see how the resulting variable changes. As they go through the model, revenues, costs, tax rates, discount rates, capital expenditures, depreciation, and other prior factors all have an impact on the net present value number. By tracing back all of these previous variables, we can change each of them by a predetermined amount and assess the effect on the resulting net present value. Due to its shape, the most vulnerable preceding variables are depicted first, in descending order of magnitude, on a graphical depiction that is frequently referred to as a tornado chart. With this information, the analyst can evaluate which crucial aspects are deterministic in the future and which are very uncertain. The uncertain important variables that drive the net present value and, thus, the decision are known as critical success drivers. For these critical success criteria, Monte Carlo simulation is an excellent fit. Because several of these critical success determinants are linked—for example, operational costs may rise in proportion to the quantity sold of a particular product, or prices may be inversely associated to quantity sold—a correlated Monte Carlo simulation may be required. The majority of the time, historical data can be used to make these relationships. When you run correlated simulations, you get a lot closer to the real-world behavior of the variables.

Real Options Problem Framing

The dilemma now is what to do after quantifying hazards in the previous stage. The risk data gathered must be transformed into actionable intelligence in some way. So what, and what do we do about it, just because risk has been estimated as such and such using Monte Carlo simulation? The solution is to apply actual options analysis to mitigate these risks, value them, and position yourself to profit from them. The act of defining the problem generates a strategic map, which is the first stage in real possibilities. Certain strategic options for each project would have been obvious based on the overall problem identification that occurred during the initial qualitative management screening phase. The strategic options could include, for example, the ability to expand, contract, abandon, switch, choose, and so on. The analyst can then choose from a list of choices to investigate further based on the identification of strategic options that exist for each project or at each stage of the project. Real options are incorporated into projects to protect against downside risks and to profit from upswings.

Real Options Valuation and Modeling

The resulting stochastic discounted cash-flow model will have a distribution of values thanks to Monte Carlo risk simulation. As a result, simulation models, analyzes, and quantifies each project's unique risks and uncertainties. As a result, the NPVs and project volatility are distributed. We assume that the underlying variable in real options is the project's future profitability, which is represented by the future cash-flow series. The results of a Monte Carlo simulation can be used to calculate the implied volatility of the future free cash flow or underlying variable. Usually, the volatility is measured as the standard deviation of the logarithmic returns on the free-cash-flow stream (other approaches include running GARCH models and using simulated coefficients of variation as proxies). Furthermore, in real options modeling, the present value of future cash flows for the base case discounted cash-flow model is used as the initial underlying asset value. Real options analysis is used to determine the strategic option values for the projects using these inputs.

Portfolio and Resource Optimization

Portfolio optimization is a step in the analysis that can be skipped. Because the projects are usually associated with one another, management should view the results as a portfolio of rolled-up projects if the analysis is done on numerous projects. Viewing them individually will not offer the actual picture. Because businesses don't just have one or two initiatives, portfolio optimization is essential. Because certain projects are interconnected, there is potential for risk hedging and diversification through a portfolio. Portfolio optimization takes all of these factors into account to build an optimal portfolio mix because firms have limited budgets, as well as time and resource constraints, while also having needs for particular overall levels of returns, risk tolerances, and so on. The research will determine the best way to allocate funds across multiple projects.

Reporting, Presentation, and Update Analysis

Until reports can be created, the analysis is not complete. Not only should the results be communicated, but so should the process. A complex black box set of analytics is transformed into transparent processes by clear, simple, and exact explanations. Management will never accept outcomes from black boxes if they don't know where the assumptions or data come from or what kind of mathematical or financial manipulation is going on. Risk analysis presupposes that the future is uncertain and that management has the authority to make mid-course corrections when these uncertainties or risks are resolved; the analysis is typically performed ahead of time and, therefore, ahead of such uncertainty and risks. As a result, if these risks are identified, the analysis should be updated to integrate the decisions made or to revise any input assumptions. Several iterations of the real options analysis should be undertaken for long-horizon projects, with future iterations being updated with the newest data and assumptions.

Understanding the processes required to complete the IRM process is critical because it reveals not only the technique itself but also how it differs from previous analyses, indicating where the traditional approach finishes and the new analytics begin.

Benefits

IRM is a great tool for improving the quality of information accessible while making decisions because it combines multiple proven strategies. When applied to the examination of potential initiatives and investments, dynamic Monte Carlo simulation depicts the risks connected with the projects in a more realistic manner than traditional methodologies. Static forecasting based on assumptions and past performance provides a restricted view of a project's potential outcomes. Decision-makers can acquire a more full understanding of the project's uncertainty by running thousands of simulations or more while altering the variables within realistic possibilities. Increasing the amount of relevant and correct information available to managers will increase the quality of the leadership team's decisions.

IRM takes a methodical strategy to deal with AI investments. Following the eight phases is a simple procedure that aids in the quantitative decision-making process. While the functions within each phase can be sophisticated and require additional training, the overall process is straightforward and simple to follow. Because the IRM approach is fully defined, it may be integrated into existing procedures without requiring a complete reengineering. IRM will use data from existing approaches and expand it to improve the scope of a project's evaluation. The true possibilities were quantified, and the outcome diverged from what was expected. The systemic design of IRM allows different members or teams to finish the process without having to re-collect data and start from the beginning. Analysts should be able to continue the procedure from any point in the approach after completing IRM training.

Real options analysis provides managers with the probability of certain project results, allowing them to select the best way to proceed with a project. Real options were offered not only at the start of the program, with three different routes in which the program may go, but also at each stage of the chosen strategy. By drafting a contract that allows an organization to modify its course of action as more information becomes available, the corporation can reduce losses from failing programs while maximizing gains from initiatives that are succeeding or showing promise. Fortunately, many viable possibilities are already ubiquitous in DoD buys. Contracts are frequently canceled by the government due to changes in budgetary policy, inability to satisfy requirements, or other factors. Including other genuine choices in contracts isn’t an entirely new concept.

The use of common units to make strategic decisions about a system’s value is a core component of the IRM methodology. Leadership can see a statistical range reflecting the potential value of a project by incorporating KVA values into the static and dynamic IRM models. The present values of the genuine option strategies were calculated using the market comparable prices produced by the value analysis. The effectiveness of most other ways is determined only by the program's cost, presuming that the value is inherent owing to the needs that were produced. IRM can provide decision-makers with information on both the expenses of a proposed investment in an initiative and the value of that project in comparable units.

Challenges

While IRM is a very useful analytical tool, it does have some disadvantages. The method’s multiple techniques might be challenging to master (Housel et al., 2019). To do a full study, it is a hard process that necessitates a solid understanding of both finance and statistics. While computing tools can help with the analysis, the inputs are more involved than simply typing a few numbers into a program and receiving the results. An analyst can generate the essential information to enable decision-makers access to the proper comparison material to make an informed decision if they have a good understanding of the core principles, enough training, and the right tools (Housel et al., 2019). The amount of data gathered during statistical analysis can be overwhelming. The simulations and their conclusions appear to originate from a quantitative black box to individuals without a strong statistics background (Mun, 2016). If decision-makers don't comprehend why an analyst makes a recommendation, it’s simple to dismiss the advice and fall back on tried-and-true methods. To tackle this possible issue, create detailed and complete reports for management review, as well as knowledgeable presentations to allay worries about the unfamiliar procedures. To take advantage of actual options, they must be reviewed before a decision is made to implement any of them. When writing contracts, leadership must consider the future option to ensure that certain alternatives stay available. Some alternatives, such as expanding, can be implemented very easily by building a new project based on the first investment’s success. However, if the contract does not include relevant conditions, project managers may not have as much flexibility in abandoning the project. Vendors must be willing to accept the possibility of subcontract cancellation when they are not at fault, which may increase the cost of completing a task. Managers must perform a careful study of which prospective options may be exercised in the future before signing contracts with vendors, due to the potential increased cost associated with contracting genuine options.

IRM, like all financial forecasting, makes projections based on previous data. Decision-makers can gain more insight from predictions that incorporate current information rather than relying just on historical trends. Meteorologists, for example, compile weather forecasts from a variety of sources: Current weather conditions are monitored using Doppler radar, satellites, radiosondes (weather balloons), and automated surface-observing systems (National Oceanic and Atmospheric Administration [NOAA], 2017). The data from multiple sources is run in models based on known historical patterns for the region using numerical weather prediction (NOAA, n.d.). Knowing the present conditions is just as crucial to a meteorologist as knowing the past models (NOAA, n.d.). Similarly, the models would deliver even more precise information if the project analyst could add pertinent information that is up to the minute (or to the requisite quality). Because of previous projects with historical data, outsourcing, lowering manning and retaining the current structure all offer statistics that could be used in simulations. Despite the fact that this weakness is not exclusive to the IRM technique, executives should be aware of it in any financial forecast.

Finally, the DoD not currently reward PMs who reap the rewards of risk. The risk framework in DoD acquisitions is intended to reduce project costs and schedule overruns. DoD contracts are structured in such a way that they do not incentivize vendors or the project as a whole to improve their capabilities or performance. When a for-profit company invests in an initiative that may fail, it does so because the potential upside gain outweighs the risk of failure. For example, if an aircraft's design target is to attain 250 knots and the design threshold is 200 knots, the budget will be allocated to the threshold rather than the objective. Unless the PM is able to reallocate resources internally, the program will not be able to meet its objectives. The acquisitions process considers the cost of achieving the goal rather than the worth of the goal. Performance is rewarded in for-profit businesses, which is evaluated by revenue. The DoD's implicit surrogate for revenue is cost reductions, which has a different value than improving a project's worth. Acquisitions by the DoD, on the other hand, are only made when the negative implications have been mitigated to the maximum extent practicable. The upside risk is unimportant to the PMs; all that matters is that the program is finished on time and on budget. Although it is still important to look at how potential projects fit into the DoD's broader collection of acquisitions and current assets, the contract structure limits certain IRM portfolio optimization features.

Comparison of Key Attributes

The type of methodology to use should be determined by the nature of the project at hand, including the level of commitment required from the organization, the organization's desire to align strategic goals with the project, the methodology's predictive capability, the flexibility required, and the amount of time available. While others in the business must understand concepts in order to comprehend status reports, EVM just requires the management team to track the project's cost and schedule against the baseline because there is no pre-determined goal alignment with the organization. While the CPI and SPI can assist in estimating the ultimate cost and schedule, EVM has no true predictive potential because it is assumed that the schedule would follow the baseline regardless of historical performance volatility. In EVM, sticking to the baseline is critical, and altering requirements can substantially affect the baseline, lowering the methodology's effectiveness. For an AI project with its many unknown components and capabilities a priori, setting up, monitoring, and reporting the cost/schedule performance of each work item inside the WBS can be a time-consuming and costly operation.

To assess the value of a process or component output, KVA simply requires the KVA analyst and the process owner, who serves as the SME, supporting the requirement to match the project with an organization's productivity goals. They can model the present baseline as-is process ROK and compare it to the proposed to-be process model ROK using this approach, resulting in a straightforward forecast of the improvement between the models. Because KVA can be used with any description language that defines process outputs in common units, analysts can choose the method that is most helpful for the system in question, allowing for flexibility. This analysis may be conducted fast, with a rough assessment available in a few days. To assess how a project fits into an organization's portfolio, the project's present value (PV), and potential real possibilities, IRM requires organizational leadership, portfolio and project managers, and the analyst. IRM gives a prediction of a project's anticipated performance by analyzing and simulating alternative situations, allowing managers to build in flexibility via genuine options at the right spots within the project. Assuming that the data required for the analysis is available, the process can be done quickly.

Methodologies in AI Acquisition

As previously stated, each methodology has strengths and weaknesses that make it more appropriate for certain applications than others. The iterative nature of software development is the most difficult aspect of adopting EVM when gaining AI. To be most successful, EVM requires well-stated, specific requirements for intermediate phases. While software program outputs are well specified, the methods required to produce the software are not, causing challenges when estimating cost and schedule. EVM can adequately monitor the progress if the software is not complex or comprises well-known operations. Integrating software and hardware is also difficult with EVM since there are various elements of the program that must be merged to achieve the objectives, requiring additional debugging and recoding. When used to manage the physical production of systems or infrastructure, EVM is more efficient. It can track the cost and schedule progress of software work packages, but it's not as good at determining their worth.

Any IS system can use KVA to offer an objective, ratio-scale measure of value and cost for each core process and its subprocesses or components. Managers can then examine productivity ratios information, such as ROK and ROI, using the two factors to determine the efficiency of a process in relation to the resources utilized to create the output. This can assist the manager in deciding how to allocate resources for system updates or estimating the future value of a system that is being purchased. Managers can iterate the value of system real options analysis using simulation and other ways by combining KVA and IRM data. IRM can also use past data to evaluate risks and anticipate performance probability for metrics of potential success for programs and program components. It's a tool that can help with investment strategy and can be used to acquire any form of AI. It is not, however, intended to assist in the procurement of an AI program or in determining how to meet the program's specific criteria.

Summary

The scope, capabilities, and limitations of various AI systems are demonstrated by examining the benefits and challenges of the proposed approaches. It also aids in determining which areas and phases of the Defense Acquisition System life cycle the methodologies or components of the methodologies should be included. The following section offers suggestions based on the findings.

Conclusion

Simply put, how might certain advanced analytical decision-making processes be applied in the acquisition life cycle to supplement existing procedures to ensure a successful acquisition of AI technologies?

As previously stated, EVM is the sole program management methodology that the U.S. government requires for all DoD acquisition initiatives worth more than $20 million. Regardless of this necessity, EVM is a methodology that offers a systematic approach to IT acquisition through program management processes that can assist in keeping an acquisition program on track and below estimated cost estimates. However, there are substantial drawbacks to utilizing EVM for AI acquisitions, the most prominent of which is that it was not built to manage AI acquisitions that follow a highly iterative and volatile course. Organic AI acquisitions necessitate a high level of flexibility in order to deal with the unknowns that surface during the development process, as well as value-adding opportunities that were not anticipated. Furthermore, EVM lacks a uniform unit of value metric that would allow typical productivity metrics like ROI to be calculated. When a program's worth is determined by how closely it adheres to its initial cost and schedule projections, the program's performance may suffer in terms of output quality when intended program activities become iterative, as in the development of many AI algorithms. EVM is not designed to recognize disproportionate increases in value if an AI acquisition program is going toward cost and schedule overruns, but the ensuing value-added of the modifications to the original requirements offers disproportionate increases in value.

To address EVM's shortcomings in AI acquisitions, the methodology should be combined with KVA and IRM, which can be useful during the EVM requirements and monitoring phases by ensuring that a given AI acquisition is aligned with organizational strategy and that a baseline process model has been developed for establishing current performance prior to the acquisition of an AI system. After the AI has been obtained, a future process model that forecasts the value-added of incorporating the AI can be used to set expectations that can be tested against the baseline model. IRM can be used to anticipate the value of strategic real choices flexibility that an acquired AI might bring, allowing leadership to choose the alternatives that best meet their desired goals for AI in defense core activities.

KVA should be utilized in AI acquisitions because it gives an objective, quantitative measure of value in common units, allowing decision-makers to better comprehend and compare different strategic options based on their value and cost. Only by employing KVA to determine the value inherent in the system can AI systems be given a return on investment. PMs benefit from this information since it gives them a more full picture of the current and future systems' performance.

When obtaining AI through the Defense Acquisition System, it's also a good idea to use IRM. The risk estimates associated with the components and subcomponents of a program, in terms of potential cost overruns, value variabilities, and schedule delays, can be improved by using dynamic and stochastic uncertainty and risk-based modeling techniques to predict likely and probabilistic outcomes. Analyzing multiple real-world options in the context of the models' outputs will assist PMs in making the best decisions possible when defining the future of a program.

As is now done, PMs should only employ EVM throughout the Engineering and Manufacturing Development (EMD) phase. EVM, on the other hand, will operate best in hardware manufacturing solutions with fully mature technology prior to the program's start. EVM is not well suited for AI development because many AI acquisition efforts involve upgrading current technology and generating new software solutions to meet requirements. Nonetheless, PMs can employ a variety of agile EVM strategies to complete projects on time and on budget if the proper procedures are done when establishing the baseline. Requirements must be broken down into tiny, simply defined tasks, with risk and uncertainty elements appropriately accounted for in the timetable. Other approaches, such as KVA and IRM, should be used in conjunction with EVM to guarantee that these elements are based on verifiable measurements rather than assuming how much more time, money, and value may be required to execute complex tasks.

KVA and IRM will assist in determining the value of the various options evaluated in the analysis of alternatives (AoA) process during the Materiel Solution Analysis (MSA) phase. KVA can objectively assess the value of the current, as-is system as well as potential future systems. Then IRM can leverage other aspects like cost, value, complexity, and schedule to value the alternatives in terms of their respective parameter values. As the chosen solutions mature during the TMRR phase, a revised KVA analysis will reassess initial estimations and provide a predicted ROI that may be incorporated into an IRM risk and actual alternatives analysis for the AI solution before entering the EMD phase, if necessary.

Limitations and Future Research

This study looked into whether the various methodologies—EVM, KVA, and IRM—could be used to improve AI acquisition inside the Defense Acquisition System. Future research should look at how these approaches interact with or improve other acquisition system components. This comprises the specific procedures of JCIDS and PPBE, as well as the interactions between JCIDS, PPBE, and the Defense Acquisition System as a whole. Certain approaches, such as IRM, may be more useful when applied to the full acquisition process rather than just a part of it. Future research might also look into how these diverse methods could be utilized to acquire things that aren't related to AI or IT.

The study focused on AI as a whole, rather than individual types of AI. Future research should look into whether acquisition methods, strategies, and methodologies differ depending on the type of AI being acquired. This is particularly relevant when it comes to artificial intelligence and its subsets. Based on their complexity, intricate nature, developing technology, and amount of risk, machine learning, intelligence with a specific emphasis or field of specialty, and general or universal intelligence will likely use different ways in the acquisition process.

Another area of prospective investigation is the use of these approaches in commercial AI acquisition. The focus of this study was solely on the application of the strategies in the DoD acquisition process. Commercial entities, on the other hand, face challenges when adopting extensive or complicated AI and IT systems, especially when the technologies are used at the enterprise level. Further research may reveal whether these same techniques could be useful to private-sector decision-makers during the development, adoption, or customization of commercial AI. The hype cycle for AI and automation is on the rise, as highlighted in the literature, and the demand to buy such technology is as relevant for the private sector as it is for the DoD. In addition, the current pandemic triggered by Coronavirus Disease 2019 (COVID-19) has compelled a permanent shift in society toward permanent distant labor. Because these trends are expected to continue in the near future, more automation tools will be needed to support this workforce. As part of the Fourth Industrial Revolution and Industry 4.0, these developments could be investigated for their consequences.

Finally, this study looked at only the most promising approaches out of a wide range of options. Other program management tools, management philosophies, analytic tools, or other approaches, as well as their benefits while adopting AI, should be investigated in future research. While the approaches investigated were chosen because they are likely to enhance the process and assist EVM improvements, other systems may be more appropriate in certain phases or provide additional benefits not seen in this study.

References

Betts, K. D., & Jaep, K. R. (2017). The dawn of fully automated contract drafting: Machine learning breathes new life into a decades-old promises. *Duke Law and Technology Review*, *15*(1), 216–233.

Darken, R. (2019, October 21). *Human–machine teaming AI*. Naval Postgraduate School.

DoD. (n.d.). *PALT*. https://DoDprocurementtoolbox.com/site-pages/palt

DoD. (2003, May 12). *The defense acquisition system* (change 2, DoD Directive 5000.01). https://www.esd.whs.mil/Directives/issuances/DoDd/

DoD. (2005). *Work breakdown structures for defense materiel items* (MIL-HDBK-881A). http://www.acqnotes.com/Attachments/MIL-Hand%20Book-881.pdf

DoD. (2011, April). *Technology Readiness Assessment (TRA) guidance*. Assistant Secretary of Defense for Research and Engineering.

DoD. (2014, March 12). *Risk management framework (RMF) for DoD information technology (IT)* (change 2, DoD Instruction 8510.01). https://www.esd.whs.mil/Directives/issuances/DoDi/

DoD. (2015a, January 7). *Operation of the defense acquisition system* (change 4, DoD Instruction 5000.02). https://www.esd.whs.mil/Directives/issuances/DoDi/
DoD. (2015b, September). *DoD program manager's guide for integrating the cybersecurity risk management framework (RMF) into the system acquisition lifecycle*.
DoD. (2017, February 26). *Defense acquisition guidebook*. https://www.dau.mil/tools/dag
DoD. (2018, April 16). *Agile and earned value management: A program manager's desk guide*.
DoD. (2019a, January 18). *Earned value management implementation guide (EVMIG)*.
DoD. (2019b, February 12). *Summary of the 2018 Department of Defense artificial intelligence strategy: Harnessing AI to advance our security and prosperity*. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF
Eykholt, K., Evitimov, I., Fernandes, E., Li, B., Rahmati, A., Xia, C., & Song, D. (2018). *Robust physical-world attacks on deep learning visual classification*. Cornell University.
Gardner, H. (1993). *Multiple intelligences*. Basic Books.
Golstein, B. (2018, October 10). *A brief taxonomy of AI*. SharperAI. https://www.sharper.ai/taxonomy-ai/
Greenfield, D. (2019, June 19). *Artificial intelligence in medicine: Applications, implications, and limitations*. Harvard University. http://sitn.hms.harvard.edu/flash/2019/artificial-intelligence-in-medicine-applications-implications-and-limitations/
Gunning, D. (2017, November). *Explainable artificial intelligence*. DARPA. https://www.darpa.mil/attachments/xaiprogramupdate.pdf
Housel, T., & Bell, A. H. (2001). *Measuring and managing knowledge*. McGraw Hill.
Housel, T., & Kanevsky, V. A. (1995). Reengineering business processes: A complexity theory approach to value added. *Infor*, *33*(4), 248. https://doi.org/10.1080/03155986.1995.11732285
Housel, T., & Kanevsky, V. (2006). *Measuring the value added of management: A knowledge value added approach*. Naval Postgraduate School.
Housel, T. J., Mun, J., Jones, R., & Carlton, B. (2019a, May). *A comparative analysis of advanced methodologies to improve the acquisition of information technology in the Department of Defense for optimal risk mitigation and decision support systems to avoid cost and schedule overruns* [Paper presentation]. Sixteenth Annual Acquisition Research Symposium, Monterey, CA, United States.
Housel, T. J., Mun J., Jones, R., & Carlton, B. (2019b, October). *A comparative analysis of advanced methodologies to improve the acquisition of information technology in the Department of Defense for optimal risk mitigation and decision support systems to avoid cost and schedule overruns* (NPS-AM-20-002). Naval Postgraduate School. https://dair.nps.edu/handle/123456789/2774
Icertis. (n.d.-b). *Customers*. Retrieved October 11, 2019, from https://www.icertis.com/customers/
Icertis. (2019a). *Icertis customer profile: Mindtree*. https://www.icertis.com/customer/mindtree/
Icertis. (2019b, April 2). *Microsoft streamlined its contract management*. https://www.icertis.com/customers/microsoft-information-exchange-agreements-case-study/
Institute for Robotic Process Automation & Artificial Intelligence. (2019, November 16). *What is robotic process automation?* https://irpaai.com/what-is-robotic-process-automation/
King, A. D. (2019, November 16). *Talk to transformer*. https://talktotransformer.com/
Knight, W. (2017, April 11). *The dark secret at the heart of AI*. Technology Review. https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/
Lawgeex. (2018, February). *Comparing the performance of artificial intelligence to human lawyers in the review of standard business contracts*. Law.com. https://images.law.com/contrib/content/uploads/documents/397/5408/lawgeex.pdf
Leviathan, Y. (2018, May 8). *Google Duplex: An AI system for accomplishing real-world tasks over the phone*. Google AI Blog. https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html
Mun, J. (2015). *Readings in certified quantitative risk management (CQRM)* (3rd ed.). Thomson-Shore and ROV Press.
Mun, J. (2016a). *Real options analysis* (3rd ed.). Thomson-Shore and ROV Press.
Mun, J. (2016b, October). *Empirical cost estimation tool* [Paper presentation]. Naval Acquisitions Research Conference, Monterey, CA, United States.
Mun, J. (2019). Empirical cost estimation for U.S. Navy ships. *Universal Journal of Management*, *7*, 152–176.

Mun, J., & Anderson, M. (2021, January). Technology trust: System information impact on autonomous systems adoption in high-risk applications. *Defense Acquisition Research Journal, 28*(1), 2–39. https:// doi.org/10.22594/10.22594/dau.19-841.28.01

Mun, J., George, K., & Ledbetter, E. (2020). *Total ownership with life-cycle cost model under uncertainty for surface ships' electro-optical-infrared-sensors* [Unpublished manuscript].

Mun, J., & Housel, T. (2010). *A primer on applying Monte Carlo simulation, real options analysis, knowledge value added, forecasting, and portfolio optimization.* Calhoun.

Mun, J., Housel, T., & Wessman, M. D. (2010). PEO-IWS ACB insertion portfolio optimization. In *Proceedings of the Seventh Annual Acquisition Research Symposium* (Vol. 2, No. NPS-AM-10-069-VOL-2). Naval Postgraduate School. https://my.nps.edu/documents/105938399/108624025/NPS-AM-10-069.pdf/c71c6830-853a-448b-beac-242bea4cba8b?version=1.0

Nayak, P. (2019, October 25). *Understanding searches better than ever before*. Google. https://blog.google/products/search/search-language-understanding-bert

Parloff, R. (2016, September 28). From 2016: Why deep learning is suddenly changing your life. *Fortune.* https://fortune.com/longform/ai-artificial-intelligence-deep-machine-learning/

Raghaven, S., & Mooney, R. J. (2013). *Online inference-rule learning from natural-language extractions*. The University of Texas.

Rich, B. (2018, February 12). How AI is changing contracts. *Harvard Business Review.* https://hbr.org/2018/02/how-ai-is-changing-contracts

Sammalkorpi, S., & Teppala, J. P. (2019). *AI in procurement*. Sievo Oy.

Shanahan, P. (2018). *DoD cloud strategy*. Department of Defense.

Sievo. (2019, November 16). *AI in procurement*. https://sievo.com/resources/ai-in-procurement

Zarkadakis, G. (2019, September 11). *The rise of the conscious machines: How far should we take AI?* Science Focus. https://www.sciencefocus.com/future-technology/the-rise-of-the-conscious-machines-how-far-should-we-take-ai/

Two Gaps That Need to be Filled in Order to Trust AI in Complex Battle Scenarios

Bruce Nagy—is an applied Research Engineer at the Naval Air Warfare Center, Weapons Division, at China Lake, California. His focus is on integrating game theory and machine learning to create recommendation engines that handle complex, highly dimensional battle management scenarios. In the Navy, Bruce served as an Engineering Duty Officer, where he developed statistically-based algorithms to enhance satellite communications. He was also a technical troubleshooter for the NRO and received honors for his efforts in recovering a failing national defense program that jeopardized U.S. global security. Nagy is also driving a joint effort to establish guidelines for an appropriate level of rigor to be used during the creation of AI-enabled systems. He has received four degrees in math, science, and engineering, and he has conducted postgraduate work analyzing brain stem to muscle group neurology. [bruce.m.nagy.civ@us.navy.mil]

Abstract

In human terms, trust is earned. This paper presents an approach on how an AI-based Course of Action (COA) recommendation algorithm (CRA) can earn human trust. It introduces a nine-stage process (NSP) divided into three phases, where the first two phases close two critical logic gaps necessary to build a trustworthy CRA. The final phase involves deployment of a trusted CRA. Historical examples are presented to provide arguments on why trust needs to be earned, beyond explaining its recommendations, especially when battle complexity and opponent surprise actions are being addressed. The paper describes discussions on the effects that surprise actions had on past battles and how AI might have made a difference, but only if the degree of trust was high. To achieve this goal, the NSP introduces modeling constructs called EVEs. EVEs are key in allowing knowledge from varying sources and forms to be collected, integrated, and refined during all three phases. Using EVEs, the CRA can integrate knowledge from wargamers conducting tabletop discussions as well as operational test engineers working with actual technology during product testing. EVEs allow CRAs to be trained with a combination of theory and practice to provide more practical and accurate recommendations.

Introduction

What does trust in Artificial Intelligence (AI) mean? October 2020, Sandia National Laboratories (SNL) conducted a “Trusted Artificial Intelligence” roundtable with national leaders and industry experts. According to SNL, “AI is trusted if its output can be used in key decision making, including cases where lives may be at stake” (Sandia National Labs, 2021). In a May 26, 2021, memo outlining DoD Plans for Responsible Artificial Intelligence, Deputy Secretary of Defense Dr. Kathleen Hicks stated:

> As the DoD embraces artificial intelligence (AI), it is imperative that we adopt responsible behavior, processes, and outcomes in a manner that reflects the Department’s commitment to its ethical principles, including the protection of privacy and civil liberties. A trusted ecosystem not only enhances our military capabilities, but also builds confidence with end-users, warfighters, and the American public. (Hicks, 2021)

Trusted AI has quite a high bar to pass. Not only must AI be trusted in cases where lives may be at stake and the mission at risk, it must also have the confidence of the American public.

Trust in AI, like trust in human relationships, takes time to build. AI must first prove itself in multiple iterations of complex, realistic synthetic battle scenarios before it can be trusted in actual conflict. The American public needs to know that it’s been thoroughly tested, evaluated, and validated before it’s used to place their sons, daughters, husbands, wives, fathers, and mothers in harm's way. AI reliable performance can also be considered a system safety issue involving hazards that can embarrass the U.S. (Nagy, 2021).

The degree to which AI should be employed in a complex battle scenario is directly proportional to the degree of trust in that AI system. In other words, the greater the objective, the greater the requirement for trust. For more important decisions, where lives may be at stake, it should require an even higher level of trust, but it also depends greatly on decision-makers weighing risk versus the reward for using AI. For instance, an opponent seeking to get inside an adversary's decision cycle might leverage AI to the degree that he or she determines that the reward outweighs the risk. Likewise, an adversary seeking to disrupt an opponent's decision cycle may want to increase their perceived risk, or choose to mistrust their AI systems.

There are many sayings regarding the need to "earn trust," or the need to "build a relationship based on trust," and an AI system making potential life and death recommendations should not be an exception. This paper will provide a nine-stage process (NSP) describing an approach that creates a trustworthy COA recommendation algorithm (CRA), proven through reliable performance in various functional roles. This approach helps to ensure that decision-makers can be confident in the COA recommendations, especially when life is at stake or an adversary is actively working to create mistrust. Will the forecast of surprise attacks and/or recommendations to commit military resources be trusted? To answer this question, the NSP provides a process that allows the CRA to: (Gap 1) learn tactics and strategies through professional wargaming via tabletop discussions, and (Gap 2) analyze performance limitations and strengths with greater statistical accuracy of products (resulting from technology development/acquisition) via "live" operational testing within Test and Evaluation, Verification, Validation/Live Virtual Construct (TEVV/LVC) facilities. The paper will present why these gaps need to be filled to adequately build trust in a CRA providing critical recommendations within complex battle scenarios.

CRA Tasking in Wargames (Gap 1)

A CRA needs to be developed from a wargaming environment to capitalize on a "treasure trove" of move-to-counter-move knowledge and possibilities, such as: (1) human factors that can affect outcomes, (2) unanticipated/surprise moves changing battle results, (3) multi-domain scenarios, where joint and coalition forces are integrated to achieve a common goal (DSB, 2015), and (4) the ability to accurately interpret various qualities of intelligence/sources. A CRA needs to learn how to unravel battle complexity, including uncovering and managing "unknowns" (DSB, 2009), and still determine an optimal strategy/tactical response. Uncovering "unknowns," meaning revealing surprises in battle before they happen, is challenging. In terms of AI systems, describing "unknowns" within complex battle scenarios, as well as how they can be uncovered or countered before the event, will be reviewed when discussing Event-Verb-Event (EVE) chains and modeling of wargames. A CRA must collect move-to-counter-move knowledge and possibilities, and learn from wargaming experts in order to provide recommendations that can be trusted.

Dr. Peter Perla defines a war game as "a warfare model or simulation whose sequence of events is interactively affected by decisions made by players representing opposing sides, and whose operation does not involve the activities of actual military forces" (Perla, 1987). Perla goes on to state, "The true value of wargaming lies in its unique ability to illuminate the effect of the human factor in warfare. By their very nature, war games seek to explore precisely those messy, 'unquantifiable' questions that campaign analyses ignore. War games can help the participants discover what they don't know they don't know," (Perla, 1987). Wargames, exercises, and campaign or operations analysis are all useful tools to build AI trust. However, exercises tend to be costly endeavors with scripted timelines and campaign analysis is often bound by analytical frameworks. Only wargames "allow for the continual adjustments of strategies and tactics by both sides in response to developing results and events not seen in campaign analysis" (Perla, 1987). Only iterative wargames over time, with opposing blue and red team members, can render insights into future conflicts.

CRA should be able to use varying levels of intel reliability about the opposing side. Based on intel and performance knowledge of its own technology, it should show strategic and tactical bottlenecks, strengths and weaknesses, as well as ways to improve self-resiliency to ensure success of mission from both red and blue perspectives. Additionally, CRA should be able to adjudicate red and blue moves and countermoves. The CRA design needs to store complete wargames, with details, and then stochastically reenact the wargame to collect statistical results for analysis. It should also be able to alter the levels of intel reliability for either opponent, and through additional analysis show trends and variations. The CRA needs to provide support of blue, red, and white players in three ways:

1. Run the wargame from a blue perspective: It needs to run the wargame from a blue perspective (with related allies) based on what blue "thinks" red (and related allies) will do; but use red "true" actions and intent during the game. In other words, it's just a blue teammate experiencing, with its blue team members, how well it anticipated red actions. This is analytical assessment from a blue teammate perspective. The CRA needs to learn and share those statistical results with blue team members regarding how to prepare better for unanticipated, "out-of-the-box" surprises in battle from a blue perspective.
2. Run the wargame from a red perspective: CRA performs that same tasking as in the previous step, just from a red perspective. This is considered "playing red with fidelity and rigor" (Rielage, 2017). The CRA needs to learn and share those statistical results with red team members regarding how to prepare better for unanticipated, "out-of-the-box" surprises in battle from a red perspective.
3. Run the wargame from a white perspective: The CRA needs to adjudicate red and blue team moves. It needs to simulate "what if" scenarios. This is an analytical assessment from a white cell player perspective, knowing performance truth of blue and red teams. The CRA needs to provide the blue team with a move-by-move analysis on how intel and/or technical capability were used, skewed, hidden, or even missed, and how those decisions impacted results. The CRA needs to do this for blue and red, using the white perspective to help white team players during debriefs of games. The CRA role needs to be constantly assessing patterns from thousands of seemingly uncorrelated data to learn how to minimize impacts from unanticipated, "out-of-the-box" surprises when making recommendations to its blue and red teams for the next set of wargames. Analyzing thousands, potentially millions of data points, should be a natural CRA function to perform. These types of analyses would be directed by white cell players to discover strategic and tactical bottlenecks, strengths, and weaknesses, as well as to improve resiliency of the systems completing the various missions.

The CRA needs to be able to analyze past wargames to support comparisons of tactical structures based on intel as well as to provide options during evaluation. As more data is collected, the more the CRA is able to provide statistical evidence, segment by segment, following the NSP. It can then use this evidence to make recommendations to blue and red teams on how to be more effective in achieving their mission goals. For the white team, it would automate adjudication, saving wargamers' time and allowing for more statistically precise analysis of moves. This would start to fill the first gap by providing value to red, blue, and white players in support of their goals and work habits.

A final reason to have the CRA integrated into the wargaming environment is complexity. Does the training data sufficiently represent the deployed challenges involved with battle complexity (Nagy, 2021)? If not, how can a COA recommendation be trusted when battle complexity is experienced during operational deployment? If trust is being earned, then the CRA must make recommendations, with a high statistical likelihood of success that consider battle complexity challenges. During a complex battle engagement, when functions affecting loss of life, property, or key objectives hang in the balance, trusting an unproven/inexperienced CRA in battle, even with human oversight, seems unreasonable. Yet, the motivation for deploying a CRA includes faster reaction time, avoiding human loss, and if trained properly, greater precision in action or detail in recommendation. Therefore, the issue becomes the meticulous process of training or evolving the technology into a trustworthy CRA. The only way this training can be successfully demonstrated is by having a high percentage of successful outcomes when following the recommended COA.

Variables that need to be considered: Does training data adequately prepare a CRA to reliably perform when challenged with the complexity of battle? Will the wargames include the proper complexity? Complexity consists of four elements that when combined make something complex: adaptability, interdependence, interconnectedness, and diversity. According to the scientific definition of complexity, a problem is more complex if it has more of these characteristics (Frank, 2015). Complexity theory is command and control theory: both deal with how a widely distributed collection of numerous agents, acting individually, can nonetheless behave like a single, even purposeful entity (Schmitt, 2008). Most times in literature, the definition of battle complexity can be summarized as a situation where there are many military components, systems, and subsystems interacting for a single purpose against an equally complex opposing force. Note: the term “complicated” relates to difficulty. Dr. Bonnie Johnson and CAPT Scot Miller, U.S. Navy (ret.), both from the Naval Postgraduate School, and other research scientists have written papers and lectured expressing “unknowns” or “uncertainties” being core to complexity (Johnson, 2019; Logan, 2009).

CRA Tasking in Operational Tests (Gap 2)

A logical next step is to have the CRA move from wargaming tabletop discussions to working with actual “live” operational testing of new technology products being developed/acquired by Department of Defense (DoD) programs. It is important that the CRA learn from firsthand experience what products can and cannot do. This data can then be used to refine the moves and countermoves discussed during the wargaming exercises. This also ensures accuracy in the recommendation. When the CRA has demonstrated reliable COAs based on the guidance of wargamers, the CRA then needs to refine its knowledge using “live” data.

In support of TEVV/LVC facilities, the CRA also needs to be designed to write test scripts that can more accurately identify the strengths and weaknesses of new technology products being developed/acquired by the DoD. By analyzing how well the new DoD technology performs when challenged by the test script scenario, the CRA offers additional value to the wargamers while supporting the operational test engineers.

To be valued, it must provide an automation capability to reduce time in developing test scripts and ensure adequate coverage of requirements. It must also share the analytical and statistical knowledge gained through wargaming to support the operational test engineers in developing more tactical and strategic battle complex test scripts. Performance data can then be used for future wargaming, allowing for any tabletop corrections regarding product performance, and thereby adding additional value to professional wargamers. These types of operational test scripts produced by the CRA will satisfy two goals:

(1) To understand how well developmental AI or any new technology can handle the unexpected, i.e., surprises, in terms of performance, capability, and resiliency
(2) Allow CRA to bridge the gap between what is operationally tested and how it is used in a professional wargaming environment; this added value earns “credits” with regard to trust

Again, from a common-sense standpoint, before the CRA is deployed in an operational environment, the goal is to provide the CRA with performance results from following its created test script scenario. The performance results would include working technology supporting the operational test, as well as the product being reviewed for release. Additionally, it’s important the CRA can refine, modify, and even correct assumptions/performance data originally described in the wargaming exercises. It must learn as much firsthand knowledge about working/deployed technology being used in the operational theater as possible. It is vital to include this practical performance knowledge in the training of a CRA expected to provide trustworthy recommendations when deployed.

Bridging the gap between wargamers and operational test engineers is important to consider. How well do professional wargamers and operational testers share knowledge? Wargamers speak in terms of strategies, tactics, and outcomes. Testers speak in terms of requirements, performance capabilities, and statistical results. Do operational test script scenarios adequately or correctly reflect how wargamers’ scenarios use those technology/product assets when games are played out? The CRA, following the NSP, ensures this alignment. The paper will demonstrate how the NSP, by using EVE modeling, will align these two domains and allow the CRA to produce cohesive and trustworthy recommendations.

The Certainty of Surprise in Battle Engagements

In a 1955 news conference, President Dwight D. Eisenhower stated, “Every war is going to astonish you in the way it occurred, and in the way it is carried out” (Eisenhower Library, 2022).The current conflict in Ukraine doesn’t appear to challenge this assertion, even with six plus decades of new technology. In a 2018 Center for Strategic and International Studies (CSIS) report titled ~~*Avoiding*~~ *Coping with Surprise in Great Power Conflicts*, Mark F. Cancian concludes that “surprise is inevitable” (Cancian, 2018) but also points out four different types of surprise: strategic, technological, doctrinal and political/diplomatic. The report analyzes each type of surprise in detail, making clear that not all surprises are a result of adversary action. Doctrinal surprise, according to Cancian, “is the use of known capabilities or technologies in unexpected ways that produces powerful new effects. Doctrinal surprise can also come from the unexpected failure of our own warfighting concepts” (2018). A recent example of doctrinal surprise is the 2020 Armenian–Azerbaijan conflict in which the Azeri used armed UAS to catch the Armenians by surprise and tipped the scale of conflict in favor of Azerbaijan (Canadian Army, 2021). Trusting AI/machine learning (ML) requires these systems account not only for an adversary’s surprise but also when our systems, processes, and procedures unexpectedly fail to work as advertised.

From the CSIS article, Cancian defined surprise as “when events occur that so contravene the victim’s expectations that opponents gain a major advantage” (Cancian, 2018). This definition of surprise is too broad for an AI system designed to measure individual bits of data. In AI terms, s*urprise* may result in an opponent gaining a major advantage but the origin of surprise may come from a completely unexpected event, or the cumulative effect of many little *surprises* causing deviation beyond toleration in the AI algorithm. This paper does not address whether or not AI may eventually remove the element of surprise from warfare, although, in the authors opinion that is unlikely. Rather, this paper analyzes what is required for AI to be trusted in future conflicts with the element of surprise ever present.

Surprise is inherent in warfare—considered unbound data—this should be considered a given fact. Another given fact is that AI battle decision aids have been known to “catastrophically” fail when presented with unbounded information (Moses, 2007; Cooter, 2000). How is this problem addressed in AI? Ensure the data sets used to train the AI system accurately reflect the deployed operational state! Ergo, the need for extensive wargaming and operational testing before deployment. Filling these two gaps are not optional; they are required to ensure trust in the CRA.

Lessons Learned on why CRAs Need to Earn Trust Before Operational Deployment

History is replete with examples of the United States being surprised (DSB, 2009), including Chinese entry into the Korean War, North Vietnamese offensive during the Tet holiday, Egyptian and Syrian attacks on Israel in 1973, the fall of the Shah in 1979, the fall of the Berlin Wall in 1989, terrorist attacks on September 11, 2001 (Cancian, 2018), and the tenacity of Ukrainian civilians to stand up to a Russian attack on their homeland.

Could a CRA have predicted the December 7, 1941, attack on Pearl Harbor? If so, would leadership have trusted the prediction? Cancian (2018) points out that the attack on Pearl Harbor was predicted—the problem was a lack of trust in those predictions. It seemed implausible that the Japanese would attack knowing the likelihood of bringing the United States into the conflict. More importantly, what would be needed for leadership to change existing battle plans, dedicate resources, and spend the needed operational funds? Given the strategic location of Pearl Harbor and critical vulnerability of point of loading (POL) logistics, it is very plausible that wargamers predicted this possibility. Cancian (2018) also points out, “The United States had broken the Japanese diplomatic code (MAGIC) and therefore had extraordinary insights into Japanese thinking and intentions. Nevertheless, for a variety of reasons—tight controls over access, gaps in information, delays in transmission, confusion about meaning, preconceptions about where an attack might occur—this extraordinary trove of data was not adequate to alert U.S. forces.” If wargamers could have predicted this attack, could an AI decision aid tool have made the same prediction? Even if the prediction was not considered credible, would an optimized, trusted resiliency plan have made a difference (DSB, 2009)?

Looking at the Allied Island-hopping campaign in WW2, there are some battles where AI decision aid tools would have been surprised and probably not effective. Using AI, battle loss might have been minimized with an optimized, trusted resiliency plan. In the best case scenario, the AI decision aid tool would have brought years of both wargaming and operational testing experience to the battle commander. Could that have been used to improve Allied warfighting effectiveness in World War II, both from move, countermove recommendations, as well as resiliency plans in case of surprise events?

Explainable AI May Not Be Enough When Significant Change Is Needed

Using “what ifs” to examine the Pearl Harbor attack from the perspective of the United States (who lost the battle), a question to consider is, could a CRA following the NSP have made a difference in the outcome?

- What if the CRA had participated in running wargames and testing of technology involved with surprise attacks at 3rd Fleet/Pearl Harbor? Would it be able to identify variations in defense preparation, resiliency plans, and tactical recommendations? If so, that data could be used to explain a recommendation to commit forces for a surprise attack. Would the recommendation have been followed using this data?

- What if some of the variations included statistical likelihoods, minimal defense postures, and pattern recognition of Japanese force movements and had been forecast through the CRA's training process? How much of a difference would explaining these details have helped in getting people to prepare?
- What if the CRA, from running the wargame over and over, learned how to minimize response time to deal with a surprise attack, possibly by ensuring resiliency as defensive preparation, or provide a core counterattack that minimized impact? How many casualties could have been avoided? Yet, with this explainable data, would U.S. leadership have listened?
- What if the CRA had earned a trusted relationship with its Pearl Harbor decision-makers, maybe during professional wargaming events or at test facilities? In other words, the CRA had already impressed its users by giving reliable recommendations in wargames and/or provided analysis that created more effective use of technology to achieve mission results. Given this proven track record, would it have made a difference to the users in choosing to follow its recommendations?

An interesting point regarding this "what if" scenario is that there was a plethora of data regarding the Pearl Harbor attack, but the base cadre did not react. An overwhelming amount of data pointed to a surprise attack. Someone reviewed the data and concluded that that there was a likelihood for a surprise attack. But, no one believed the accumulated data enough to support a commitment of military resources. The lesson learned may be that explainable AI, the human reviewing the data acting as an AI equivalent, was likely not enough. Would coming from a computer have made a difference? The definitive answer is performance history!

The goal is to build a trusted relationship with the CRA. Yet, trust is earned through performance reliability, i.e., it generates a high degree of successful recommendations. It's earned by being a reliable wargaming recommendation tool that has demonstrated its ability to counter unknown-unknowns. It can also earn trust through value-added knowledge from operational testing. Following the NSP, the CRA can develop and earn a positive reputation! Consider that without this proven reputation, no matter how explainable, would a recommendation from an AI algorithm be considered reliable enough to commit sizeable amounts of military assets? The point is that a CRA needs to build its reputation based on performance to earn trust!

If this conclusion has credence, the Pearl Harbor lesson is significant. Even if the perfect AI recommendation system is developed, without past history of trust, explainable AI is not enough. Even if the AI explains its recommendation using past history/training data, for example, that the Japanese attacked the Russian's Port Arthur in China about half a decade ago, explainable AI would not be sufficient. (This history was well known at the time.) The point is that as explainable/historical as the recommendation might be, without trust based on a proven track record, the result of the attack on Pearl Harbor would likely have remained the same. Without a trust history, a military commander is not likely to commit a sizeable number of military resources based on a machine's recommendation. Additionally, a resiliency plan recommended by the CRA, even if it was perfect, may have suffered the same fate because it lacked a history of trust.

Consider the Battle of Midway from the perspective of the Japanese who lost the battle. Would an AI recommendation algorithm have made a difference in that outcome?

The Japanese knew that they had superior forces, more experienced pilots, better aircraft, and an element of surprise. In a wargame that calculates the odds of winning, the

Japanese likely determined that they would emerge victorious nearly 100% of the time. As a result, the “unknown-unknowns” for the Japanese significantly affected the outcome of the battle, i.e., they were wrong. They did not account for the Americans breaking their code, which is a surprise in technology capability. Their calculations did not account for the heroic and nearly suicidal efforts of many naval aviators—a human factors surprise. Japan assumed that rearming and refueling, dangerous operational tasks, would always be handled with utmost care, meaning taking time, but they favored speed, another human factor surprise. Again, consider the “what if” list and whether a CRA could accumulate enough knowledge to uncover these unknowns. If so, would the recommendations have been trusted? If these unknowns could have been uncovered, would the CRA have been believed regarding recommendations to counter these surprise events? Again, this is a relationship issue, earned through performance reliability, i.e., to earn trust, a high percentage of successful recommendations need to be made in wargames. Yet, to make accurate recommendations, the CRA also needs to understand the performance strengths and limitations of products/technology from firsthand knowledge, learned during operational tests.

As described in the Pearl Harbor Attack and the Battle of Midway, interpretation of the intelligence and human factors plays a major role in action and reaction, move and countermove, eventually leading to a final outcome. Maybe the surprises could not have been predicted, but what if the CRA provided a resilience plan that effectively countered the impact of the surprise? As a lesson learned, intel and human factors need to be included in the training of the CRA and the evaluation of its reliable recommendations, both a proactive counter and/or an effective resilience strategy.

Trust to Overcome Hubris May Be the Best Approach

If not considered, human factors can be a surprise element during a wargame. Hubris can adversely affect a rational decision, and trust might be the only human factor that can create needed clarity. How much trust is enough to overcome hubris?

What are the lessons learned when hubris plays a role in making decisions? This question is important to consider because the United States is considered a “superpower.” Can the hubris make a CRA recommendation even harder to accept? Does it raise the bar regarding how much trust needs to be earned to overcome hubris for the recommendation to be accepted? Can hubris become a weakness, impacting battle outcomes? Was hubris a major factor in the lack of reaction to overwhelming data stating the Japanese was about to attack Pearl Harbor? Did the Japanese demonstrate hubris during the Battle of Midway attack?

A potential example of hubris might be in an interpretation of the Battle of Nagorno–Karabakh War from the perspective of the Armenians who lost the battle. This is purely conjecture regarding attitude and must be emphasized that this discussion is being provided as an example only. This interpretation may be false, but will be used to emphasize a point regarding the potential that hubris may make it more difficult to trust an AI system providing CRA recommendations. The conjecture is that the Armenians assumed a weaker opponent and although intel stated a buildup of capability across the border, hubris of their past success overruled their caution. The result is that Azerbaijan actually proved themselves during battle to be a peer adversary. This was a surprise to the Armenians.

Armenian confidence was based on a history of success with Azerbaijan, considered “known-knowns” (past history). Azerbaijan confidence was based on increased “known-unknowns” (assumptions about improvements). The Armenians won the last war and thought they would win the next (assumptions). They were not prepared for Azerbaijan’s improved battle capability (surprises). On the other hand, Azerbaijan learned from the last war. They increased their technology and military training by linking with winning Russian technology and strategies.

As a result, they were able to significantly alter the Armenian's expected outcome of the battle. Azerbaijan's had Russian's Snowdome defense and Armenian's did not anticipate its effectiveness. Was this poor intel or hubris? This was a technology surprise factor. The effectiveness of UAS (used in good weather) and tank artillery (used in bad weather) severely reduced Armenian capability—additionally, use of the Israeli Harop (UAS) to provide both surveillance and kinetics was effectively used, another Azerbaijan technology surprise.

Would a CRA, trusted through proven recommendation performance through wargaming and operational testing, have enough earned reputation to overcome any potential Armenian hubris? Like in the U.S. example where intel pointed to a Pearly Harbor attack, intel data was not sufficient. With regard to the United States, does its status as a superpower cause hubris among its military leadership, and would a CRA with a proven track record in wargaming and operational testing (e.g., understanding the performance effects of Snowdome or Harop) have made a difference? Would AI explainability of the data used to make the recommendation be enough? How much trust would have been needed, along with explainable data, to convince Armenians that they needed to better prepare?

Avoid Designing a CRA to Earn Limited Trust

Another human factor to consider is a willingness to die for one's belief. This has been true with suicide bombing. Suicide bombing is a surprise tactic that has occurred in Arabic wars, as well as during World War II. Japanese Kamikaze bombing was completely unanticipated. From the perspective of the United States, could this have been foreseen? Could this sacrifice have been anticipated by a CRA? The first kamikaze pilot to drive his airplane into a WWII warship likely would have been a significant departure from predicted norms, and a surprise to any modern-day AI system. Could the CRAs have dealt with kamikaze attacks? It is possible a modern-day AI system could have analyzed Bushido code (Anya, 2013) to recognize that Japanese culture placed a great deal in sacrificing life for honor and from that made a correlation to the possibility of future kamikaze attacks. The solution is obviously valuable, but does it earn trust?

This correlation relies on AI programmers to input the Bushido code to support a kamikaze prediction. This type of training data is considered bias. The challenge is that the variance between data sets would be very poor, meaning that the data would not support other cultural relationships from other countries, e.g., suicide bombers following a different religious code. As a lesson learned, it is important that CRA training avoid this bias limitation. This paper is not recommending excluding this approach, but ensuring the trust is not dependent on it when dealing with opponent surprises. If it is, then the CRA would be limited to Japanese actions associated with the Bushido code. The trust would be earned within this realm, but not others. As will be described, CRA needs to be a generalized, structured approach to deal with the wide variety of opponent surprises. This is important if the CRA is to be trusted, i.e., bias and variance should be balanced.

As an alternative, generalized approach, a kamikaze attack may not have been anticipated, but the CRA may have considered how to deal with specific types of impacts from opponent moves and countermoves. This is the benefit from using EVEs. The CRA may also have developed a resiliency plan based on assuming opponent success, another benefit from EVE modeling, thereby minimizing the effects of the opponent impact. Recommended readiness and resiliency, especially coming from a trustworthy CRA, is a proven defense against the unknown. The key is having the CRA understand vulnerability points and to make trustworthy recommendations for countermoves that include resiliency. Training of the CRA to understand vulnerability and recommend the needed response is provided using EVE Chains.

The Power of EVE Chains for CRA Development

Data needs to be collected during wargaming and operational tests as part of the training process. EVE chains are designed to replicate any type of action or exchange of actions (Nagy, 2021; Nagy, 2022). In wargaming, EVEs can model moves and countermoves based on the world state. An EVE segment can represent a specific move, countermove interaction, capturing each wargaming interaction into EVE segments for reuse. In operational test, it can represent a sequence of actions required for the product under test to perform, including evasion and other forms of counteractions. The EVE model consists of events, state variable changes resulting from verb execution. State variables comprise an event. The verb modifies certain state variable, thereby creating a resulting event. Here are some common terms used with regard to the EVE modeling:

- Event: All or part of a world state at a specific timeframe – the world state consists of all enabler and influencer state variables involved with game play. In this TAWC construct, events consume no time during game play.
- Verb: An action available to the Enabler and Influencer that changes the world state and consumes time on the game board. Verbs can be functionally represented by certified meta-models (Nagy, 2022), ML algorithms, or polynomials. Note: the combination of EVEs with meta-models allows for lightweight, low-processing power systems proven by the Battle Readiness, Engagement Management (BREM) prototype project (Nagy, 2022).
- Enabler: An asset, a "piece" within the game, that has specific Verbs (or actions) that when performed can affect the world state, e.g., Enabler Verbs can counter the negative effects of entity influencer actions and counter obstructions; or enabler verbs can take advantage of an obstruction that supports mission success. Note: Depending on perspective, Enablers can be blue or red game pieces. Enablers are only represented by Entities.
- Influencer: There are environmental and entity influencers. Environmental influencers consist of moveable and immovable obstructions, as well as weather conditions above and below. Entity influencers have Verbs that when performed can negatively affect one or more Enabler Entity state variables, as well as moveable obstacles in a way that causes mission failure. Note: Depending on perspective, Influencer Entities can be blue or red game pieces.

Training data, for both wargames and operational tests, are EVE chains that can be collected and statistically analyzed. CRA put together EVE chains as recommendations, i.e., a high statistical percentage of successful outcomes, involving actions (verbs) based on input state variables (events that lead to mission success. EVE segments are created from data collected from wargames, i.e., moves and countermoves, as well as product test results, i.e., test scripts demonstrating performance of required moves and countermoves. From wargaming and test operations, EVE segments, moves and countermoves, can be defined and refined from the same pool, supporting greater model accuracy. Greater model accuracy means opportunities for greater recommendation accuracy.

To reiterate, a necessary ingredient in the NSP, or any COA development approach, is to earn enough trust, meaning an extremely high percentage of successful EVE chains that included responses to surprise opponent actions, that if the CRA predicted a surprise event, the user would follow its recommendation resulting in a commitment of military forces. Yet, forecasting an EVE chain that is a surprise, meaning never identified within a wargame or operational scenario, is a difficult challenge.

It should be noted that professional wargames are not about learning to predict the future, nor validate friendly or enemy courses of action, i.e., EVE chains. As Perla (1987) stated, it's about "illumination" and "exploration." For the CRA learning process, wargames provide "illumination" and "exploration" of causality. It provides the medium for causal analytics that support the development of EVE sequences. Those EVE sequences lead to outcomes, based on wargame results. When those sequences are statistically analyzed, then the outcomes can be associated with a likelihood of success. If basic causal analytics can be learned from wargames, then the CRA, playing as all three team colors, can develop more statistics and EVE segments than by playing against itself. It's this accumulation of EVE segments that will support COA's being prepared to deal with surprises (Nagy, 2022).

The CRA is designed to statistically forecast outcomes based on pattern matching EVE segments accumulated during human play or self-play. It tracks actions to EVE sequences and segments as a pattern matching approach. By pattern matching EVE segments, a statistical forecast of an outcome can be produced. As an example, a person (1) wakes up every day at the same time, (2) makes coffee, (3) gets dressed in professional clothes, (4) gets in a car, and (5) goes to work. This is noticed 75 times, and 25 times the person went for gas. If the first four were provided, statistics would be obvious with regard to going to work or getting gas. But would this result be believed?

What if one of four actions was missing from the input. How would the CRA respond with a recommendation for an outcome? A given state is used to determine an action, but when the state is inaccurate, the challenge is for the CRA to still perform reliably. This is a common "garbage in, garbage out" problem and is considered a Bayesian approach to prediction (Adamski, 2019). The NSP requires the CRA to be designed to use ML generalization to deal with this issue. A surprise is when the person gets a ride from a friend. Maybe the car is in the shop. How does the CRA handle this surprise?

From an EVE chain focus, unknown-unknowns events have two parts:

(1) Part 1 is identifying "what" (one of more state variables in the Event) will be impacted that will prevent Verb (or action) from executing, e.g., blow up fuel depot to prevent planes from refueling, or destroy runway to prevent planes from taking off. Each variable can represent a one for available or zero for not available. This is called a binary EVE chain analysis. The opponent wishes to create zeros, thereby preventing any actions. A single zero in the binary EVE chain analysis can impact the success of a mission.

(2) Part 2 is anticipating "how" the opponent will cause the event (one or more state variables) to be impacted, set to zero, e.g., the process that blew up the depot or destroyed the runway. Was it a suicide bomber, a well-placed bomb, or something completely unanticipated?

From an EVE chain focus, there are two responses to counter the attack:

(1) Part 1 is determining the opponent's "how" and then an appropriate counter (with an EVE chain sequence) to ensure the state variables remain one, thereby ensuring that the EVE chain/sequence can continue to achieve mission success.

(2) Part 2 involves creating a contingency EVE chain/sequence that considers that a counter is ineffective, meaning the state variable to set to zero. And EVE chain must show resilience to the impact, i.e., an alternative Verb that maintains a successful mission outcome.

In an EVE chain, there may be thousands of state variables. The CRA, through the process of learning from wargaming and operational testing, has the time to "crunch" through thousands of state variables that might be targeted. It can assess which state variables have the highest impact, optimal counter response and resiliency plan. It can also assess which variables are least attacked but have highest impact, thereby analyzing and sharing unlikely but impactful vulnerability points. Notice that the CRA may not be able to predict how variables will be attacked, but can anticipate likelihood based on impact and counter strategies, including resiliency plans.

For EVE chains associated with AI systems, battle complexity is defined as a situation that can be described by a series of events, i.e., EVE chains, caused by actions between opposing participants, where the outcomes can be significantly affected by factors categorized as: (1) "known-knowns" (facts), (2) "known-unknowns" (assumptions) (3) "unknown-knowns" (absent data) and (4) "unknown-unknowns" (surprises; Nagy, 2021).

- "Known-knowns" (facts)—factors that participants depend on as "fact" to win the engagement; these can include own participant's ISR and C2 technical capabilities, geo-spatial, temporal situational awareness, interoperability, EW effects, human skills, tactical actions and strategy pros/cons. These are EVE chains from data collected from wargames and operational tests.
- "Known-unknowns" (assumptions)—factors that each participant needs to "assume" about variations (of the facts) regarding battle conditions, these can include the third-party involvement, weather forecast, IO, ISR and C2 effectiveness, kinetic and non-kinetic effectiveness, opponent's attack surfaces and related vulnerabilities, heroism and initiative on all sides, opponent's priorities, and difficulty in overcoming manmade and natural obstructions. These are assumed variations in EVE chains from data collected from wargames and operational tests.
- "Unknown-knowns" (absent-data)—factors that cause a participant to be "absent of data," sometimes decision critical info; these factors can include human mistakes, sensor failures, and communication issues. These are missing state variables in EVE chains.
- "Unknown-unknowns" (surprises)—factors that will "surprise" participants during the engagement; these include unforeseen technology and anything not anticipated in the previous three categories. These are EVE chains that have not been identified in any wargame or operational test. The NSP will describe how these EVE chains are addressed using generalization (Stage 9, Phase III of the NSP approach).

In a complex battle, where surprises are certain, how do you provide an algorithm with training data, i.e., EVE sequences, to handle surprises when those surprise are unknown? Consider how the CRA is being developed to support this need through wargames and product testing that include unknowns, i.e., degrees of unbound data. This is the reason why these two gaps need to be filled. This is also the reason why the CRA must play against itself, i.e., self-play, to accumulate EVE segments from a variety of moves and countermoves. The CRA can also determine a resiliency plan, another set of EVE segments, when a state variable is least likely to be attacked but has the greatest impact toward mission success remains flipped. Can enough self-play reduce the number of unbound data issues, meaning surprise events? This needs to be determined, but collecting data from wargaming does help.

Even if "surprise" is a given, it is believed that the number of surprises can be reduced through the wargaming effort, thereby reducing the opportunity for unbound issues. This is another important reason why the CRA must learn from wargamers by capturing those EVE segments. In a 1960 speech to the U.S. Naval War College (USNWC), Admiral Nimitz remarked, "The war with Japan had been re-enacted in the game rooms here (USNWC) by so many people and in so many different ways that nothing that happened during the war was a surprise—absolutely nothing except the kamikaze tactics toward the end [of] the war; we had not visualized those"(Nimitz, 1965). For more than a decade during the interwar period, wargamers at the Naval War College had war-gamed every aspect of a potential conflict with Japan and identified nearly every contingency, and yet the war with Japan still brought surprises.

The “earning trust” challenge will always involve the ability to prepare the CRA to handle unbound data issues, i.e., surprises the opponent might unveil during an engagement. Collecting data from wargaming plays a needed role to minimize those surprises, thereby reducing unbound issues. If surprises, like a kamikaze attack, do occur, the CRA needs to be prepared to provide resilient solutions along with counter solutions, both represented by EVE chains.

There should always be a concern that a surprise might cause the data input to go beyond that algorithm’s variation limits. To address this concern, there must be an approach to ensure oversight against these unbounded solutions (Miller, 2021) and ensure that the CRA is ready to provide resiliency recommendations as an alternative. Guardrails and gates are proven approaches for AI algorithms. Unbound data by its inherent definition means that confidence in the performance/behavior of the AI model cannot be predicted and therefore cannot be trusted. In order to represent a realistic operational set of training data, complexity of the deployed environment needs to be considered; resiliency planning must be immediately available. Again, the CRA attempts to address these considerations through its wargaming and product testing. Given surprise is a given, therefore unbound data is a given, wargaming and operational testing that includes resiliency is a needed ingredient for the CRA to earn trust. From these two environments, EVE segments can be collected and the CRA can be trained to recompose to meet variations in mission challenges. The process of training is the NSP.

NSP in Developing a Trustworthy CRA

The NSP is based on using EVEs to connect all parts of the learning process described in each stage using a common model. Details in Figure 1 show the accumulation of EVEs in each of the three phases. This is how to ensure CRAs make trustworthy recommendations, enabling decision-makers to be confident when life is on the line and a commitment of large military resources is needed. Using EVES, the three phases are connected through nine stages. In all three phases, shown in Figure 1, the CRA results in creating tactical or operational plans. Following the nine stages over three phases of the NSP, a trustworthy CRA can be created. Data collection using EVE modeling bridges knowledge between these wargaming and test domains, creating and refining improved recommendations as preparation for the CRA to be deployed. As a needed result of NSP using EVEs, wargamer and operational tester benefit from increased automation and statistical analysis. This motivates the users to continue to use the CRA in their domains, establishing a value-added approach for all involved.

It is also important to note that the NSP involves the training process to earn trust. Stage 1 of Phase I can occur in parallel with the CRA core development. When Stage 1 and CRA core development is complete, then Stage 2 and on can occur. The CRA must have its core development complete using EVE chains or similar modeling structures associated with world state variables. NSP is based on using EVE chains. An example of a CRA core design using EVE chains is provided in a paper by Bruce Nagy presented at the SPIE conference on Defense and Commercial Sensing (Nagy, 2022).

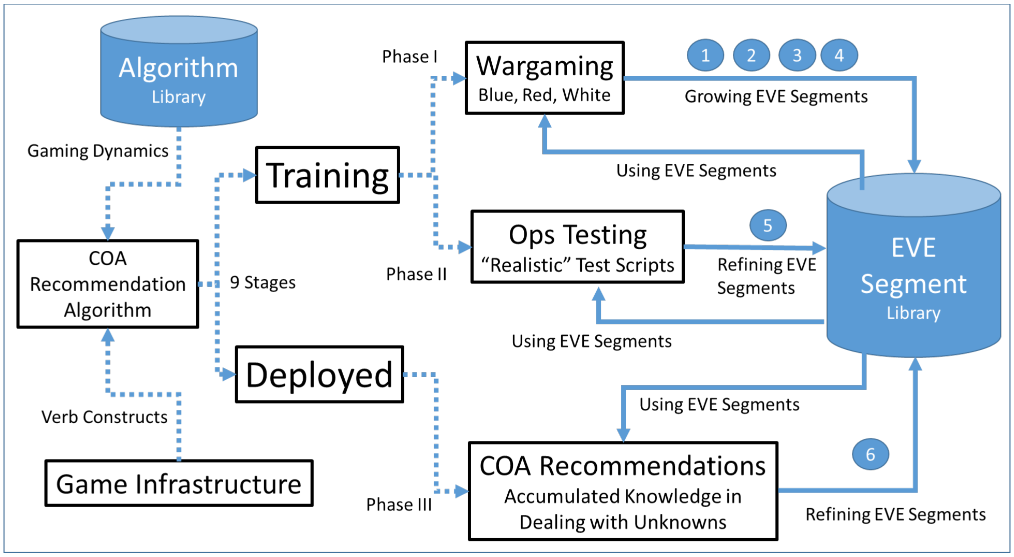


Figure 1. NSP Overview

Phase I Wargaming

The *Wargaming Mode* focuses on supporting professional wargamers. The process involves data collection from professional wargamers using validated performance capabilities of assets and technologies used in games. In this phase, the CRA performs as a wargaming analysis tool acting as a wargaming team member for red, blue, and white teams to support professional wargaming institutes in better analyzing and understanding the effects of intelligence quality, strategy, and tactical outcomes within various wargaming scenarios. There are seven stages within this phase. The first three stages of Phase I are shown in Table 1.

Stage 1 focuses on developing algorithms that move and track game pieces, i.e., assets, on the world board game. It also includes automating various adjudication processes for the wargame users. This is a prerequisite stage and must be done in advance with the focus on developing ancillary algorithm used by the CRA during game play, while also supporting automation needs of wargamers. Although Stage 1 is listed in the wargaming phase, it must also include statistical automation tools that will be used to support the TEVV/LVC facilities. Additionally, this stage establishes all the background information needed to inform the verbs and events in the EVE structure. For instance, if the verb is *move*, and it involves an aircraft entity, stage 1 captures all the performance parameters. In other words, game pieces and moves are automated for war game activity.

Stage 2 determines how the game board is initially set up and what its end goals are. It collects user data that determines what they would consider the beginning states and end states (derived from the commander's intent) for various missions. It sets the stage for the wargame, including placement of assets around the world, their state of readiness, and what goals need to be accomplished. Stage 2 captures the variety of missions, both for blue and red teams. This "current" to "end goal" states can also be entered in "real time," before the game begins. Data entry can be manual or automated about world state for the initial/first event and related state variables, as well as the last/final event, i.e., what the world state needs to "look like" when a mission is concluded. This final/last event supports the commander's guidance translated into world state variables. Notice that when state variables change based on actions, this represents EVE chains (Nagy, 2022). It is not possible to develop credible CRAs if the beginning and end states are not adequately defined. In this stage, performance bounds are also defined, providing a landscape involved with the game board.

Stages 3 and 4 involve running the CRA using these two previous stages, Stage 1 for automated game piece movement and Stage 2 for game piece placement. This is necessary if the CRA is to optimally determine the best moves and countermoves from each team perspective. Remember that the CRA takes on all team colors involved with game play.

Table 1. Developing CRA Segments 1–3

Phase I Wargaming - Segment 1:	**Phase I Wargaming - Segment 2** (for Blue as Enabler):	**Phase I Wargaming - Segment 3** (for Blue as Enabler):

• Create Verb Infrastructure
  • EVE Ontology
  • Verb Table
  • Verb Binary Codes
  • Verb Hex Code
• Create Algorithm Library
  • Geometry for Global Movement Dynamics
  • Optimization using Learning Rate of Attributes
  • Statistical Significance Analysis
• Create EVE Segment Library (with Sections)
  • Manageable Obstructions
  • Unmanageable Obstructions
  • Enabler Mission Actions
  • Influencer Actions
  • Influencer Counter Actions
★ Establishing EVE Segment Library
	• Define Mission Constraints
	  • Blue Mission Criteria
	  • Blue Performance Area
	  • Blue Environmental Influencers
	    • Immovable Obstacles
	    • Moveable Obstacles
	    • Weather and Other Conditions
	  • Blue Entities
	    • Blue Actual Performance Specs
	    • Blue Allie Estimated Performance Specs
	  • Red Influencer Entities
	    • Red Estimated Performance Specs
	    • Red Allie Estimated Performance Specs
		• Develop Ideal EVE Chain
		  • Movement Dynamics
		    • Performance Area
		    • Immovable Obstacle
		    • Environmental Conditions
		  • Mission Achievement
		    • Tree Trunk and Branches
		      • Verb Stack
		      • EVE Stack (Values)
		      • Binary and Hex Code EVE Stack
		  • EVE Segment Library by Move, Move (Results)
		    • Store Optimal Strategy
		    • Store Statistical Measured Results
		1 Growing EVE Segment Library

Stage 3 is having the CRA develop an optimal strategy and tactics for achieving the end goal state defined in Stage 2 for red and blue teams, as well as their allies. This stage views an ideal world, where opponents and environment influencers are not a factor. It states that if the board did not have opposing pieces or obstacles that it could overcome, what would be the optimal moves to achieve results, i.e., end states. This might generate many solutions that can be analyzed based on team priorities defined in terms of what is considered mission success. From this analysis, the CRA selects “best” candidate(s) with their movement domain route(s) to achieve mission success when there are no opposing/opponent entities. Obstacles may be involved, but limited to those obstructions that cannot be modified, i.e., an immovable landscape.

Stage 4 is having the CRA develop an optimal strategy and tactics for achieving end goal state when opponents forces within a movable and movable landscape, game board. It focuses on having the CRA playout various scenarios between blue and red forces. It attempts to select the optimal game piece candidate(s) with their movement route(s) to achieve mission success where there are opposing/opponent entities attempting to thwart actions. The CRA now has the ability to “go through” environmental/obstructions, if this benefits the users’ end state goals and priorities.

Stage 5 is a repeat of stages 2, 3, and 4 but for the opponent. Remember that each team only has limited knowledge, based on the quality of intel about the other player. In other words, the stages are repeated for both the blue team (with allies) and the red team (with allies) to determine their optimal strategies against each other, not knowing “truth” of the other players’ capabilities. Details associated with stages 4 and 5 are shown in Table 2. By completing Stage

4 and 5, best candidates or combinations of game pieces for each opposing team are identified to play out in a non-ideal environment, i.e., opposing/opponent entities and within moveable/manageable and unmovable/unmanageable environmental conditions and obstructions.

Table 2. Developing CRA Segments 4 and 5

Phase I Wargaming - Segment 4	**Phase I Wargaming - Segment 5**

(for Blue as Enabler):
• Develop Non-Ideal (Challenged) EVE Chain
  • Movement Dynamics
    • Performance Area
    • Environmental Influencers
      • Immovable Obstacles
      • Moveable Obstacles
      • Weather and Other Conditions
    • Entity Influencers
      • Blue and Allie Actual Performance Specs
      • Red and Allie Estimated Performance Specs
  • Mission Achievement
    • Tree Trunk and Branches
      • Verb Stack
      • EVE Stack (Values)
      • Binary and Hex Code EVE Stack
    • Counter Moves
      • Verb Stack
      • EVE Stack (Values)
      • Binary and Hex Code EVE Stack
  • EVE Segment Library by Move and Counter Move
    • Store Optimal Strategy
    • Store Statistical Measured Results
(2) Growing EVE Segment Library
	(for Red as Enabler):
	• Repeat Segment 2 to 4 for Red Team
	  • Segment 2′ (for Red as Enabler):
	    • Define Mission Constraints
	      • Red Mission Criteria
	      • Red Performance Area
	      • Red Environmental Influencers
	      • Red Entities
	      • Blue Influencer Entities
	  • Segment 3′ (for Red as Enabler):
	    • Develop Ideal EVE Chain
	      • Movement Dynamics
	      • Mission Achievement
	      • EVE Segment Library by Move and Counter Move
	  • Segment 4′ (for Red as Enabler):
	    • Develop Non-Ideal (Challenged) EVE Chain
	      • Movement Dynamics
	      • Mission Achievement
	      • EVE Segment Library by Move and Counter Move
	(3) Growing EVE Segment Library

Stage 6 involves having the CRA perform the adjudication process involved with a wargame. This means that the “white cell” runs the wargame with complete knowledge of both sides, red and blue. Their tactics and strategies were based on perception and interpretation from intel sources within the wargame construct. The CRA uses “truth” about capabilities and intent on each side to assess the actual outcome for each side in achieving mission success, given the reality of each side’s tactics and strategies. It can then run “what if” scenarios that include variations in performance and intel quality, EVE segment by segment to find optimal outcomes for each side. In game theory, this is finding either the Pure Strategy Nash Equilibrium (PSNE) or the Mixed Strategy Nash Equilibrium (MSNE). These “what if” solutions contribute to various points on a Pareto Analysis chart, i.e., a four-square readiness matrix described in segment 9.

Stage 7 compares the original results from Stages 4 and 5 to Stage 6 modifications, truth at stage 4 and truth at stage 5, comparing perception of the opponent based on intel to the actual truth of the opponent’s capabilities. Included in this comparison are the “what if” results. The process is running through each chain sequence, EVE segment by segment, and tracking how often there was an attempt to flip each state variable to zero within the binary EVE segments. The highest number becomes the most likely candidate of vulnerability and the lowest number, the least, given wargaming trends. This stage can be executed/run in the background or in advance of any wargames, as long as Stage 1 and 2 have previously been completed.

Stage 6 and 7 are shown in Table 3. These stages become a significant learning process for all involved, users and CRA, in identifying how to optimally deal with unknowns, specifically bits that were not flipped and why they were not flipped. Is there a way to create a strategy or tactic that would ensure that a mistake in assumptions has minimal effect on mission outcome? This is what the CRA is being designed to investigate and is unique from other algorithms. From a wargaming, adjudication perspective, the solution can be used for wargaming analysis and adjudication, identifying how and when to adjudicate, and providing unknown-unknown challenges.

Table 3. Developing CRA Segments 6 and 7

Phase I Wargaming - Segment 6:	Phase I Wargaming - Segment 7:

• Adjudicate War Game as White Cell
  • Run Monte Carlo Wargame with Existing Assumptions of Blue and Red EVE chains against each other
    • Blue EVE Chain using Assumptions based on Red Intel about Capability
    • Red EVE Chain using Assumptions based on Blue Intel about Capability
    • Statistics by Segment
      • Store for Enabler (Blue) with Influencer (Red) Assumptions
      • Store for Enabler (Red) with Influencer (Blue) Assumptions
  • Run Monte Carlo Wargame with "Truth" of Blue and Red EVE chains against each other
    • Blue EVE Chain using Red "Truth" about Capability
    • Red EVE Chain using Blue "Truth" about Capability
    • Store Statistics by Segment
	• Review Lessons Learned
	  • Comparison of Delta's Assumption vs Truth
	  • Connect Delta's to Statistical Results (Answers Why)
	  • Develop Optimal Solutions based on "Truth" from both sides (EVE Segments with Statistical Results)
	  • Store in EVE segment database by Move, Counter Move (Results from Blue and Red)
	    • Optimal Strategy
	    • Non-Optimal Strategy
	    • Statistical Measured Improvement
	
	4 Growing EVE Segment Library

Phase II T&E

After Stage 2 is complete, Phase II begins the CRA evolvement of earning trust. The T&E Phase focuses on supporting TEVV/LVC facilities. The process involves data refinement from testing technology products and systems using validated performance capabilities of assets and technologies used during testing. In this phase, the CRA performs as a testing analysis tool, a modification of its wargaming capability developed in Phase I. In Stage 8, per Table 4, the CRA is engineered to create rigorous test scripts, while refining EVE segments to better represent "realistic" performance capabilities, results, and limitations. There is only one stage within this phase.

The CRA is now ready for the final upgrade in becoming a recommendation algorithm to generate nominal and stress level test scripts. This is a refinement and validation process from the wargaming EVE segments. Using TEVV/LVC facilities, the EVE segments represent complex environments, linked to live systems or six degree of freedom systems. The CRA algorithm will adjust based on the performance of all products represented within the environment.

Table 4. Developing CRA Segments 8 and 9

Phase II T&E - Segment 8:	Phase III Deployed Operations - Segment 9:

• Test Script Generator
  • Import Mission Requirements for Test and other Segment 2 Data
  • Mix and Match EVE Segments to create
    • Optimal Solution: (1) EVE Tree with Causal Why, and (2) Why Statistically based on “Truth” of Influencer
    • Nominal Solution: (1) EVE Tree with Causal Why, and (2) Why Statistically based on “Truth” of Influencer
    • Stressed Solution: (1) EVE Tree with Causal Why, and (2) Why Statistically based on “Truth” of Influencer
  • Based on testing, modify EVE segments used to support measured results, including variations in statistics
  • Store Data Changes from Test Results in EVE segment database by Move, Counter Move (Results from Blue and Red)
(5) Refining EVE Segment Library
	• COA Recommendation Engine
	  • Import Mission Parameters and other Segment 2 Data
	  • Allow User Preference
	  • Mix and Match EVE Segments to create a Pareto Chart associated with BRE Matrix
	    • Solution given Assumed Intel Truth and Use Preference
	    • Solution given Variations in Intel Assumptions based on Wargaming
	  • Provide Solution that encompasses as many points on the “Green” segment of the Matrix
	    • Not optimal for a single point
	    • Best compromised solution for encompassed points
	  • Store Data Changes from “Live” Operational Results in EVE segment database by Move, Counter Move (Results from Blue and Red)
	(6) Refining EVE Segment Library

While developing test scripts and accumulating knowledge, the CRA must be engineered to collect state variables least attacked but with highest impact and not identified in wargames through state variable by state variable investigation. If found, this is considered a paradigm shift, i.e., unknown-unknowns, to support wargamers.

In support of its operational testers, the CRA needs to provide three types of test scripts. Each test script can have subscripts identifying where to change the testing conditions and scenario to support the three types of tests. The three types of test are: (1) nominal performance, (2) product performance under attack and a demonstration of an effective counter, and (3) product performance under attack, not effectively countered, therefore requiring resiliency for the product under review to examine the products limitations. All data is collected and shared with wargamers.

At this point, Phase I and Phase II are being executed simultaneously. Only after both wargamers and test engineers agree will the CRA be allowed to move into Phase III.

Phase III Operational Mode

The operational mode focuses on deploying the CRA to support assets in the field needing to make tactical and battle management decisions. The CRA learning process continues to involve data refinement from operational exercises using live data from assets and technologies. The CRA is an evolvement of the two previous phases. The CRA is now designed to provide trustworthy recommendations that will also ensure opponent generated surprise issues have minimal effect on the outcome of a mission. There is one stage within this phase.

Stage 9, as described in Table 4, represents the final CRA development stage and a graduation to live operational support, i.e., CRA being deployed. From the previous phases, EVE segments have been developed and refined, now available by the CRA when needed. In Phase I, the understanding of intel quality associated with EVE segment selection was analyzed. Additionally, the EVE segments supporting complex environments were created. This resulted in a validation of war gaming complexity and EVE solutions that are statistically significant, meaning mission impacts are unique for each solution. It should also be noted that since EVE segments were developed from professional wargamers using validated technology performance data garnered from real games, EVE segments replicated "actual" technology/asset capability. In Phase II, the next level of validation of the performance capabilities of products under test or within the test environment is established to support "firsthand" refinement of the EVE segments to ensure they represented "realism." With both complexity and realism validated, along with the CRA's understanding of intel quality effects of decisions, the CRA is now ready to develop battle readiness, engagement, and management support in the form of recommendations that have statistical significance.

Using the Pareto chart analysis approach, Figure 2, the CRA identifies a single EVE tree solution, again combining EVE segments using an AI/ML algorithm trained earlier, that supports as many point variations in the green zone as possible.

> Pareto front is a set of nondominated solutions, being chosen as optimal, if no objective can be improved without sacrificing at least one other objective. On the other hand, a solution x* is referred to as dominated by another solution x if, and only if, x is equally good or better than x* with respect to all objectives. (www.igi-global.com, n.d.)

The green zone is defined with user thresholds for probability of mission success from the Monte Carlo simulations. This is the optimal solution given a wider variety of influencer actions. The EVE tree representing these group of points in the green zone of the Pareto Chart is what is recommended to minimize effects of influencer variations and EVE tree weak points. There are two types of recommendations provided:

- Recommendation Type 1. Nominal EVE tree solution that includes as many points in the green zone as possible.
- Recommendation Type 2. Resilient EVE tree solution that supports ability to withstand a state variable flipped to zero but still support a successful mission. This resilient EVE solution must also be able to include as many points in the green zone as possible. This second type assumes unknown-unknowns occur, and because it was a surprise, successfully flipped a bit. The recommendation ensures continued success because of its solution resilience.

The "why" is provided to support explainable AI, not just in the causal relationships but also how those causal relationship caused statistical results, as collected in previous stages. The outcomes (plotted points) are known wargaming results, with variations of intelligence, depicting the Red Force's ability regarding what they do (state variable flipped), how they do it (EVE tree), and when the Blue Force is attacked. Thresholds, from one color region on the Pareto chart to the next, are determined by values calculated using the discrete correlation approach. Green area indicates that the value was considered a successful mission result when adjudicated, i.e., being above the threshold of what is considered a successful mission. Yellow areas indicate large variations regarding success and failure associated with the threshold that could not be resolved, therefore having outcomes that are uncertain. Upper, right yellow region indicates bias towards Red Force likelihood of success. Lower, left yellow region indicated bias towards Blue Force likelihood of success.

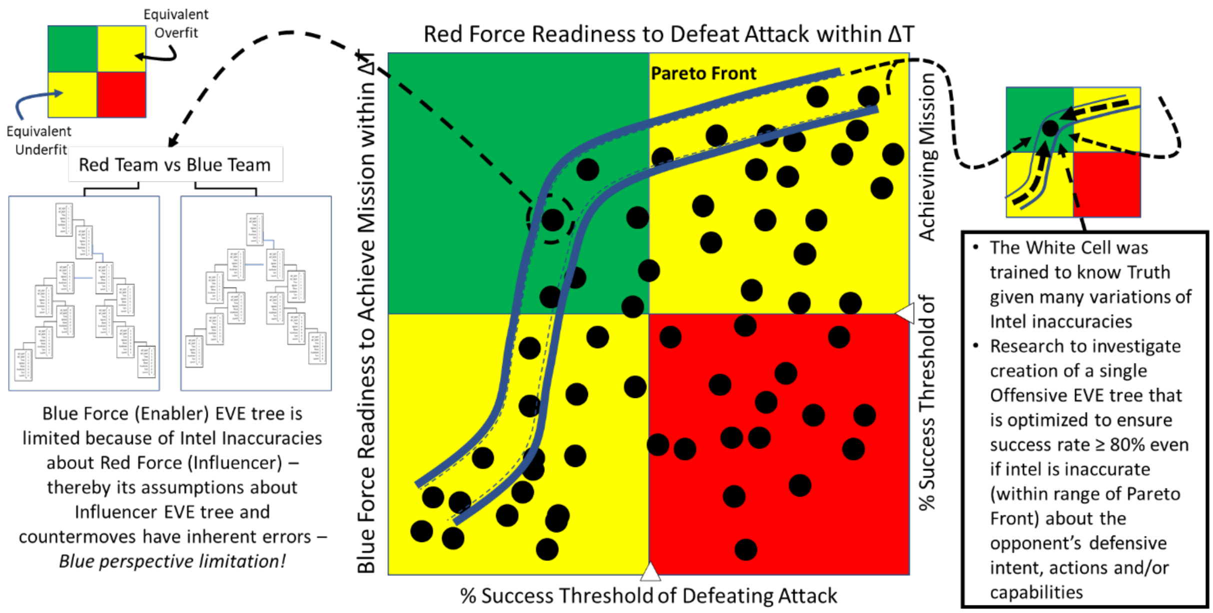


Figure 2. EVE Tree That Optimizes Ability to Succeed Independent of Opponent Strategy

As stated earlier, an "offense" action can include defensive tactics described in the EVE tree (Nagy, 2021; Nagy, 2022), and defensive actions can include offensive tactics, again described in the corresponding EVE tree. Consider how machine learning systems generalize. The CRA takes in instances of training data (points on the plot) to learn, through feedback, how to correctly determine the meaning of the input. The result is that the CRA is designed to handle variations from the original training data and still determine the meaning.

The "What's In It For Me" (WIFM) Human Factor

If users are to participate with using the CRA, there needs to be a significant return on investment for wargamers and operational test engineers, or why would they be motivated to change or do something different?

To support the motivation of professional wargamers, the goal is to automate their existing tool suite. This means moves and countermoves can be more easily entered and analyzed with significantly greater statistical precision. Likewise, since CRA following NSP can repeat the entire wargame in detail, action by action using EVE chain events, it can play out the statistical variations for "what if" analysis. It can determine confidence factors by EVE segment, meaning move and countermoves can be closely examined, even analyzing how quality levels of intel can impact strategy and tactics. Through its automation, CRA can reduce the time to set up and implement that wargame, focusing more on content and less on administration. It can create an ontology that allows for wargamers to more easily share perceptions and joint actions.

To support the motivation of operational test engineers, the CRA, following the NSP, can achieve near- and long-term goals associated with more effective analysis and testing. It can support near-term goals defined in the National Security Commission of Artificial Intelligence Final Report 2021 by providing automated decision support for constrained test scenarios that are challenged with creating "realistic" battle engagement test scripts and real/synthetic environments for autonomous systems, including manned and unmanned teaming. The CRA can be used to standardize existing TEVV/LVC facilities by providing a cost effective, common simulation environment. It will provide more accurate analysis of the strengths and weaknesses of the product under test, support resilience, and provide statistically explainable test script scenarios.

As a long-term goal defined in the National Security Commission of Artificial Intelligence Final Report 2021, the CRA will eventually be able to test an autonomous system, or a system of autonomous systems, designed to dynamically learn and adapt during a manned and unmanned teaming operation. The CRA will provide real-time decision support and course of action recommendations and auto-generated scripts. This will allow TEVV/LVC facilities to accurately replicate synthetic environments requiring open-world simulations to adequately test adaptable, autonomous platforms required to perform a wide range of joint and coalition enhanced missions.

Conclusions

The paper recommends that for a CRA to gain trust from its human users, it must be designed to fill two gaps in its training and evolvement process before being operationally deployed:

(1) Wargaming Gap (1): The CRA must learn how to provide successful recommendations during wargaming that involves complex battle scenarios that include unanticipated, "out of the box" surprises by the opposing force, or even when poor intel quality or ability to receive accurate status from its own assets are experienced.
(2) Operational Testing Gap (2): The CRA must learn how to create test scripts in support of VVTE/LVC facilities by providing requirement coverage, but also create tests that help in analyzing performance using complex battle scenarios, that include unanticipated, "out of the box" surprises by the opposing force, or even when poor intel quality or ability to receive accurate status from its own assets are experienced.

If these two specific gaps were generalized as an architype for AI development, it would be to: (1) Have the AI learn from Subject Matter Experts (SMEs), where its learning can be continually tested/validated, thereby proving performance and (2) have the AI be involved with "real" technology, learning from firsthand experience what systems can and cannot do, where its learning can be continually tested/validated, thereby proving performance. A final key aspect to using this archetype is ensuring that any human involved with the training of the AI receive value, i.e., his or her WIFM factor is also filled during the process.

These are key aspects related to both critical learning gaps (DSB, 2009) that must be filled in any CRA before being deployed to ensure trust is earned. These critical gaps are addressed by the NSP using EVE chains, and must be filled to adequately prepare the CRA for "realistic" experiences during operational deployment. During this training process, the CRA must demonstrate learned knowledge to wargamers and operational test engineers in worst case conditions, as described.

Both training gaps (1) and (2) indicate a need to work with wargaming and operational test engineers to produce battle scenarios that can help anticipate the unexpected and design an optimal response into the training data. This training must include resiliency plans for when the surprise encounter by the opponent is successful. Human oversight is still a necessity for a CRA when unwanted loss of life or property is in jeopardy, but by filling these two gaps when designing the CRA, trust will be earned through reliable performance, demonstrating the ability to deal with size and complexity of a combat situation.

Following the NSP, the CRA is designed to motivate three types of customers for continued use of the AI product. For professional wargaming, the CRA automates part of the arduous appraisal process, and provides improved analytical results, that includes causal factors. It can uniquely support wargamers with three analyses of the wargame red and blue based on intel received on each opposing side, and white knowing "truth." The CRA will be able to reenact entire wargames to statistically analyze strategies and tactics, showing bottlenecks, strengths, and weaknesses, as well as needs to improve resiliency. It can alter the intel, simulate the entire war game again, and show what ifs, trends, and variations. The AI system can learn and share those statistical results regarding how to prepare better for unanticipated, "out-of-the-box" surprises in battle from a blue perspective.

For testers, this NSP-created CRA provides test threads that enable evaluators to consider all possible uses of a particular technology in anticipated and unanticipated, but possible scenarios. It could share the analytical and statistical knowledge gained through wargaming to support the operational test engineers in developing more tactically and strategically "realistic" test scripts. The CRA would provide an automation capability to reduce time and effort in developing test scripts and ensure adequate coverage of requirements.

For operators, the CRA becomes a trustworthy tool, enabling auto generation and comparison of viable COAs, with causal, explainable factors using EVEs. It can provide COAs that can minimize red effects when limited intel is available. Further, the CRA may infer red intent and identify possible unknown unknowns, which can reduce the number of tactical surprises blue might face.

The CRA, following the nine-stage approach, has the ability to deliver new ideas on what red might do dramatically increase the blue planner and decision-maker's mental models on possible future outcomes. Moreover, the modeling of the EVE chains and related recall of EVE segments enable very rapid re-planning and generate new ways to think about and achieve operational resilience. Explaining a recommendation or action is different from developing a relationship of trust that the recommendation or action will achieve the desired result. This paper concludes that if there is a desire to minimize human involvement in complex battle scenarios (for example to improve reaction time or avoid human loss), then the AI must be able to handle the unexpected. This means the AI must be trained to handle the unexpected.

Again, the NSP is based on achieving trust through relationship with a CRA well before operational deployment. This ability to earn trust from reliable performance with wargamers and testers fills the two gaps. Additionally, the need to introduce unanticipated/surprises associated with state variables during wargaming and operational testing will enhance the ability of U.S. armed forces to prepare and overcome, resulting in less fatality, operational cost, and escalation. (DSB 2015) By using EVE chains with the three phases containing nine segments, integrating theory and practicality, it presents the potential of changing the outcome of future conflicts through COA recommendations that optimally counter unanticipated, out-of-the-box surprises by the opponent and handle complex scenarios.

References

Adamski, M., & Pence, S. (2019). Thriving in uncertainty from predictive-to-probability-based assessments. *Military Review,* March–April.

Canadian Army Intelligence Regiment, Land Force Intelligence Centre. (2021). Combined arms lessons, learned from the Nagorno-Karabakh War. Canadian Department of National Defense.

Cancian, M. (2018). ~~*Avoiding*~~ *Coping with surprise in great power conflicts.* Center for Strategic and International Studies. https://www.csis.org/analysis/coping-surprise-great-power-conflicts

Cooter, R. D. (2000). Three effects of social norms on law: Expression, deterrence, and internalization. *Oregon Law Review, 79*, 1–22.

Defense Science Board. (2009). Capability Surprise, Volume I: Main Report. DSB Capability Surprise. Vol I. https://dsb.cto.mil/reports/2000s/ADA506396.pdf

Defense Science Board. (2009). Capability Surprise, Volume II: Supporting Papers. Vol II. https://dsb.cto.mil/reports/2010s/ADA513074.pdf
DSB Strategic Surprise Final Board. (2015). DSB Summer Study Report on Strategic Surprise. https://dsb.cto.mil/reports/2010s/2014_DSB_Strategic_Surprise.pdf
Eisenhower, D. D. (2022). Quotes. Presidential Library, Museum and Boyhood Home. https://www.eisenhowerlibrary.gov/eisenhowers/quotes
Frank, A. (2015). Complexity, psychology, and modern war. *Small Wars Journal.*
Hicks, K. (2021). Implementing responsible artificial intelligence in the Department of Defense [Memorandum]. https://media.defense.gov/2021/May/27/2002730593/-1/-1/0/implementing-responsible-artificial-intelligence-in-the-department-of-defense.pdf
Miller, S., & Nagy, B. (2021). Interdependence analysis for artificial intelligence system safety. *Proceedings of the 18th Annual Acquisition Research Symposium.*
Nagy, B. (2021). Functional hazard analysis and subsystem hazard analysis of artificial intelligence/machine learning functions within a sandbox program. *Proceedings of the 18th Annual Acquisition Research Symposium.*
Nagy, B. (2021). *Tips for applying artificial intelligence to battle complexity* [Presentation]. Naval Applications for Machine Learning Symposium, San Diego, CA.
Nagy, B. (2021). *Using event-verb-event (EVE) constructs to train algorithms to recommend a complex mix of tactical actions that can be statistically analyzed* [Presentation]. Naval Applications for Machine Learning Symposium, San Diego, CA.
Nagy, B. (2022). *Battle Readiness Engagement Management (BREM) prototype: A wargaming tool to analyze "realistic," complex kill chains* [Presentation]. National Fire Control Symposium.
Nagy, B. (2022). *Deploying meta-models in warfighter laptops to drive "real-time," on-site course-of-action recommendations that solve complex kill chain solutions* [Presentation]. National Fire Control Symposium
Nagy, B. (2022). *Event-verb-event (EVE) constructs to allow machine learned systems to solve complex kill chain problems* [Presentation]. National Fire Control Symposium.
Nagy, B. (2022). *Fourteen tips to increase confidence in the performance of artificial intelligence (AI)/machine learning (ML) functions during the five stages of development* [Presentation]. National Fire Control Symposium.
Nagy, B. (2022). *Using game theory and machine learning to provide white cell automation support in the adjudication of war game* [Paper presentation]. SPIE Conference on Defense and Commercial Sensing.
Nimitz, C. (1965). Chester W. Nimitz letter to Charles L. Melson. Naval Historical Collection Archives. https://usnwcarchives.org/repositories/2/archival_objects/39995
Perla, P. (1987). War games, analyses, and exercises. *Naval War College Review.* https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=4335&context=nwc-review
Rielage, D. (2017). War gaming must get red right. *U.S. Naval Institute Proceedings, 143/1/1,367*. https://www.usni.org/magazines/proceedings/2017/january/war-gaming-must-get-red-right
Schmitt, J. (2008). Command and (out of) control: The military implications of complexity theory. http://www.dodccrp.org/html4/bibliography/comch09.html
Johnson, B. (2019). Artificial intelligence—An enabler of Naval Tactical Decision Superiority. *Association for the Advancement of Artificial Intelligence.* https://doi.org/10.1609/aimag.v40i1.2852
Logan, D. (2009). Known knowns, known unknowns, unknown unknowns and the propagation of scientific enquiry. *Journal of Experimental Botany*, *60*(3), 712–714. https://doi.org/10.1093/jxb/erp043
Moses, L. B. (2007). Recurring dilemmas: The law's race to keep up with technological change. *Illinois Journal of Law, Technology, and Policy, 2*, 239–285.
Silva, A. (2013). Unit Five: Kamikaze Pilots and Shinto. *Cultural Anthropology.* https://anyasilva89.wordpress.com/2013/05/14/unit-five-kamikaze-pilots-and-shinto/

PANEL 18. CONTRACTING STRATEGIES THAT IMPROVE OUTCOMES

Thursday, May 12, 2022	
11:05 a.m. – 12:20 p.m.	**Chair: John Tenaglia,** Principal Director, Defense Pricing and Contracting at United States Department of Defense ***Middle Tier Acquisitions and Innovation*** John Kamp, George Washington University Amir Etemadi, George Washington University ***Assessing Policy Changes on the Cost of Husbanding Services for Navy Ships*** Margaret Hauser, Naval Postgraduate School Geraldo Ferrer, Naval Postgraduate School Robert Mortlock, Naval Postgraduate School ***Federal Contracting: Senior Leaders Should Use Leading Companies' Key Practices to Improve Performance*** Timothy J. DiNapoli, US Government Accountability Office Nathan Tranquilli, US Government Accountability Office Guisseli Reyes-Turnell, US Government Accountability Office

John Tenaglia—serves as the Principal Director, Defense Pricing and Contracting (DPC) within the Office of the Secretary of Defense (OSD), U.S. Department of Defense (DoD). He is responsible for the pricing and contracting policy matters that relate to $300+ billion in annual contract obligations. As the functional leader of DoD's 30,000+ contracting and purchasing professionals, he is responsible for their workforce career development. He serves as the principal advisor to the Under Secretary of Defense for Acquisition and Sustainment (USD(A&S)) on acquisition and procurement negotiation strategies for all major weapon systems, and major automated information systems programs. He is responsible for assisting the USD(A&S) in the implementation of the Department's operational priorities, including business reform initiatives to reduce operational costs and speed delivery to the warfighter. He oversees business enterprise initiatives related to pricing, domestic, international, and contingency contract policy for DoD, and associated e-business solutions; and DoD procurement regulations. Mr. Tenaglia also served as Acting Principal Deputy Assistant Secretary of Defense for Acquisition from June 2021 to February 2022.

Prior to his current assignment, Mr. Tenaglia served as the Head of the Contracting Activity (HCA) for the Defense Health Agency (DHA). In this capacity, Mr. Tenaglia was responsible for the award and administration of all contracts to maintain a medically ready force and a ready medical force, with annual obligations valued at over $13 billion. Having previously served in DPC under its former name, Defense Procurement and Acquisition Policy (DPAP), from 2008 to 2017, Mr. Tenaglia last served as Deputy Director for Contract Policy and International Contracting. In that role, Mr. Tenaglia was responsible for developing and improving innovative procurement policies for the Department of Defense by advancing the most efficient means to acquire supplies and services for the nation's warfighters.

Mr. Tenaglia was commissioned in the United States Air Force in 1988 and retired from active duty in 2008. His Air Force tours of duty include: Minuteman III Missile Launch Officer; Contracts Manager and Executive Officer to the Director of Contracting for Aeronautical Systems Center; Contracting Officer and Contracting Policy Staff Officer at the National Reconnaissance Office; Commander of the 81st Contracting Squadron; Contracting Staff Officer in the Office of the Assistant Secretary of the Air Force for Acquisition (Contracting); Legislative Liaison in the Office of the Secretary of the Air Force, Office of Legislative Liaison; and Special Assistant for Legislative Affairs in the Office of the Assistant Secretary of Defense for Legislative Affairs.

Mr. Tenaglia earned a Bachelor of Arts degree in Political Science at the University of South Florida in Tampa, Florida. He earned a Master of Science degree in Health Services Administration from Central Michigan University and a Juris Doctor (JD) degree from Catholic University's Columbus School of Law. He has been admitted to practice law in the Commonwealth of Virginia since 2012. Mr. Tenaglia has been a member of the Senior Executive Service since 2014.

Middle Tier Acquisitions and Innovation

Dr. Amir Etemadi—Amir Etemadi received a Doctor of Philosophy degree in electrical engineering from the University of Toronto, Toronto, Canada. He is an Assistant Professor in the School of Engineering and Applied Science in the Department of Electrical and Computer Engineering and the Department of Engineering Management and Systems Engineering, and teaches masters- and doctorate-level courses in electrical engineering and engineering management. He is the Principal Investigator on several current research projects and has supervised over 50 masters and doctoral candidates for degrees in both electrical engineering and engineering management. [etemadi@gwu.edu]

Dr. John Kamp—John Kamp received a Doctor of Engineering degree in engineering management from the George Washington University. Kamp teaches masters-level courses in project and program management and is a faculty Graduate Research Advisor in the School of Engineering and Applied Sciences. He is a retired Naval Submarine Officer with extensive experience in research and development and program management. His research interests include engineering management, maritime systems, and acquisition system research. Kamp is a Fellow in the Royal Institution of Naval Architects and a member of several professional associations. [jckamp2018@gwu.edu]

Abstract

How can we tell if policy innovations such as Middle Tier Acquisitions are working as intended? This research uses publicly-released data consisting of budget submissions, program-related reporting, and contemporaneous press releases to describe how the services are using Middle Tier Acquisition authorities to accelerate system innovation. Project schedule durations and intervals between significant events are used as indicators of significant schedule innovations. Middle Tier Acquisition programs have development times like other acquisition programs, but are much faster than other acquisition processes in going from initiation to development start and from design review to fielding of a prototype or capability.

Research Issue Statement: This research examines how Middle Tier Acquisition policy innovations affected acquisition system schedule performance relative to traditional major defense acquisition programs.

Research Results Statement: This research provides quantitative assessments of the effects of Middle Tier Acquisition policy innovations on project strategies and schedules.

Keywords: Middle Tier Acquisition, Defense acquisition, innovation

Introduction

This paper reports results from research considering three specific statutory changes intended to speed delivery of new capabilities and products: modular development, Agile development, and Middle Tier Acquisitions. Major defense acquisition programs (MDAPs)[19] take about eight years to proceed from program initiation to an initial operational capability, which is longer than adversaries need to create new problems for operational military forces.

Research Scope

The research applies to Department of Defense (DoD) acquisition program innovations, including g modular development, Agile development, and middle tier acquisition (MTA) programs, and specifically excludes programs intended to acquire services or Defense business

[19] See 10 U.S.C. 2430 for an explicit MDAP definition (10 USC 2430, 2021).

systems. This research included acquisition policy and management changes enacted in the 2016, 2017, and 2018 National Defense Authorization Acts (NDAAs) and the DoD and service guidance, governance, and execution strategies implementing these changes. The research findings may not be valid for other system commodity types such as ships or ground vehicles or for acquisition practices outside the considered set of innovations.

Research Questions and Objectives

RQ1. What programmatic attributes differentiate Major Capability Acquisitions and Middle Tier Acquisitions?

RQ2. What programmatic attributes differentiate Middle Tier Acquisitions from other rapid acquisition approaches?

RQ3. What programmatic attributes differentiate Middle Tier Acquisitions and commercial New Product Development?

Research Objective: to examine how public policy innovations directly related to DoD rapid acquisition strategies affected program performance and achieved intended policy outcomes.

Background/Literature Review

Innovation Definition and Measurement

The Organization for Economic Cooperation and Development (OECD) defines innovation as “a new or improved product or process (or combination thereof) that differs significantly from the unit’s previous products or processes and that has been made available to potential users (product) or brought into use by the unit (process)”(OECD, 2019).

Patent grants are a common measure of innovation (OECD, 2019). Table 1 summarizes patents from government contracts and grants between 2000 and 2013.

Table 2. Patent Percentages from Government Awards (2000–2013)

Type award	DoD	DOE	Other
Contract	15%	12%	2%
Grant	9%	6%	57%

Based on data from (de Rassenfosse et al., 2019).

Most patents arising from government-funded research come from research grants. If patent grants are measuring innovation, the above results suggest that DoD programs in particular and government contracts in general are not innovative (de Rassenfosse et al., 2019).

Fagerberg considered National Innovation systems and showed that while they have different structures and dynamics, technological innovation is essential to economic growth, (Fagerberg, 2017). Caiazza noted governments may act to improve innovation diffusion, and identified supply-side, demand-side, and general barriers preventing diffusion from the innovator to the adopter (Caiazza, 2016). General innovation diffusion barriers are often cultural, legal, or economic barriers (Caiazza, 2016). In principle, few statutory barriers exist to DoD innovation. In practice, departmental and programmatic risk aversion (Lopez, 2021) and lack of urgency (Flournoy & Lyons, 2016) act as general barriers.

Supply-Side Barriers to DoD Innovation

Four broad trends in federal procurement suggest some supply-side barriers to innovation: Federal procurement spending, a shift in procurements from products to services, an increased contracted workforce, and geographic spend concentration (Taylor, 2019). Federal procurement spending, while growing overall, was increasingly a lower percentage overall of the federal budget; federal procurements are increasingly shifting from goods to services; an increasing reliance on contracted workforce; the increase in federal procurements in the District of Columbia-Maryland-Virginia region[20]; and the close interactions of firms with offices in the DMV region with Congress and the federal procurement system (Taylor, 2019). These trends may disrupt some existing suppliers of defense-unique products, suggesting an expanding market opportunity to not only create new products, but to change the associated acquisition processes. In particular, defense research and development is a service (General Services Administration, 2020), and subject to different parts of the Federal Acquisition Regulations (FAR) than products[21]; non-FAR authorities such as other transaction agreements (10 U.S.C. 2371, 1993) or us commercial-type acquisition methods and approaches such as Procurements for Experimental Purposes (10 USC 2373, 2015) may be used to acquire research and development services.

Lockhart advocated for open communications within the DoD and between the DoD and suppliers to improve innovation performance (Lockhart, 2018). This incentivizes the supply side to create a market for technology innovations and future sales. The Defense Innovation Unit is a different but complementary effort, outside the normal acquisition community[22], and provides market access and non-dilutive capital for non-traditional defense contractors (*DIU*, 2020). It is focused on transitioning commercial advanced technologies in specific domains to the DoD, and uses an extension of other transaction agreement authorities (NDAA, 2015, sec. 815) to fund development and transition (*DIU*, 2020).

Demand-Side Barriers to DoD Innovation

Defense demand-side innovation barriers include business processes and culture. In 2014, the Defense Business Board analyzed core DoD acquisition processes, and estimated the overhead costs of current processes, potential savings, and general recommendations on goals and processes for business process improvements (Defense Business Board, 2015). The Advisory Panel on Streamlining and Codifying Acquisition Regulations[23] provided extensive recommendations intended to accelerate acquisition processes by leveraging commercial marketplaces and processes, simplifying acquisition regulations, changing resource allocation processes, and improving the acquisition workforce (Drabkin et al., 2016)[24].

A key demand-side issue is being able to efficiently discover new innovations. Fleming and Sorenson treat invention as a complex search over technology domains, and found that the local search space size and interdependence are the most significant predictors of successful

---

[20] Turkina et al. cite such geographic proximity or density as a factor in creating innovation clusters (Turkina et al., 2019).

[21] Research and development contracting regulations are in FAR part 35, acquisition exploratory and development contracting regulations are in FAR part 34 (General Services Administration, 2019).

[22] The Defense Innovation Unit has offices in Silicon Valley, Boston, Austin, and Chicago, and reports to the Undersecretary of Defense for Research and Engineering (USD(R&E)).

[23] Created by Congress, this temporary Panel was also known as the Section 809 Panel, in reference to the National Defense Authorization Act section that created the panel (NDAA, 2015),

[24] The DoD has implemented less than half the recommendations.

search (Fleming & Sorenson, 2001). Such search traditionally required engaged expertise to discover new opportunities; these searches were typically the domain of the government research and development communities. In 2015, the defense procurement and acquisition community comprised about 20% of the total civilian workforce; the Defense Business Board recommended overall workforce reductions and retention of non-specific expertise, freeing resources to buy more systems (Defense Business Board, 2015), but likely increasing demand-side barriers, as fewer defense personnel would be aware of and in a position to adopt new innovations.

Demand-side barriers are also related to resource availability and liquidity. Congress created several funding processes designed to accelerate technology transitions from non-traditional performers and are posted in the Defense Innovation Marketplace (OUSD[R&E], 2020). For example, in 2011 Congress created the DoD Rapid Innovation Fund to accelerate small business technology transition to the DoD (NDAA, 2011). It was managed by staff within the Office of the Undersecretary of Defense for Research and Engineering (OUSD[R&E]), and structured to complement DoD small business innovative research programs by providing transition funding to move technology into operational use or to an acquisition program within 24 months (OUSD[R&E], 2020)[25]. Figure 1 summarizes small business innovative research data from 2010–2020.

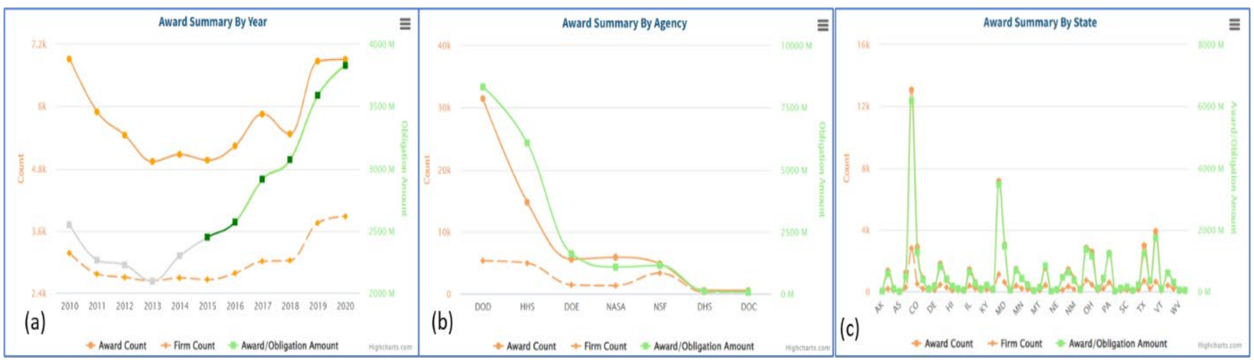


Figure 10. Small Business Innovative Research Summary, 2010–2020.
(sbir.gov. n.d.)

Figure 1a shows overall spending trends were roughly constant[26]. The DoD awarded most Small Business Innovative Research awards (Figure 1b), with most awards clustered in a few states, notably California, Massachusetts, Virginia, and Maryland (Figure 1c). Bresler noted that while the DoD has a long history of funding small, innovative companies, it generally does not provide follow-on business growth beyond initial sponsorship, reducing incentives for these companies to invest and remain in the defense market (Bresler, 2018). The conclusion is that small business research provides significant exposure to new ideas, but inefficient transition of innovations to product.

Kendall advocated structural and policy changes to control costs and to incentivize industry and government to adopt strategies such as increased use of prototypes and open

[25] Congress did not appropriate funding for this activity in 2020 or later years.
[26] The average per award increase between 2010 and 2020 was about 3% per year.

system architectures[27] (Kendall, 2014). Such recommendations are both supply-side and demand-side, as suppliers are incentivized to create new products and buyers have increased exposure to new ideas and incentives to discover new products. Policies can also create disincentives. For example, DoD major automated information systems programs were required to report significant schedule growth to Congress (Cha, 2016), creating a strong incentive for schedule adherence[28], which tends to suppress seller innovation[29], and reduces incentives for opportunistic behaviors (Schoeni, 2018).

In 2016, Congress enacted Middle Tier Acquisition (MTA) processes enabling the Department of Defense (DoD) to prototype and field new capabilities within two to five years of approval (NDAA, 2015). Four statutory changes set the foundations for accelerating new capability development: 1) explicitly setting an objective duration; 2) providing explicit authority allowing service acquisition executives to bypass traditional requirements and acquisition processes[30]; 3) revising funding approval thresholds, authorities, and applicability criteria[31]; and 4) allowing direct transition to production under specific conditions[32]. Following Fagerberg, the schedule constraint provides a *demand* for new innovations, and the revised authorities provide *institutional* and *financial* capability to execute (Fagerberg, 2017). In 2018, Congress authorized a DoD Agile Pilot program[33] (NDAA, 2017). These innovations acted to *reduce innovation barriers*.

New Product Development and Time to Market

The commercial new product development process may inhibit supply for two reasons: profitability outside government procurement and significant process factors for commercial new product development. Braha and Bar-Yam modeled commercial new product development as a complex network. They found that such networks are responsive to design status changes, processes are bounded in their ability to process new inputs, and they hypothesized that information flows tend to follow system architectures (Braha & Bar-Yam, 2007).

Markham et al. looked at informal activities in new product development. Informal interactions between three roles—champion, sponsor, and gatekeeper—precede formal new product development and represent significant research and business activity (Markham et al., 2010). Breakdowns in these interactions create gaps between research and formal development called the "Valley of Death"; Bonnin Roca and O'Sullivan thought a major cause of this gap was a lack of investment to take a proof of concept to a prototype or commercialization, and pointed to regulatory uncertainty and technology immaturity as causes of this reluctance (Bonnin Roca & O'Sullivan, 2020). This gap can occur within an organization; Dean et al. note that

---

[27] Cost control was to improve buying power; the strategy changes were to reduce technical barriers to innovation.
[28] Note that in 2015 MAIS programs had mean cycle times of about 3 years (Kendall, 2016)
[29] Schedule adherence pressure will incentivize using more products that are either in-use or commercial products.
[30] Such as establishing direct-reporting program managers for these rapid acquisition programs (NDAA, 2015).
[31] Section 815 approval authorities were modified to allow "The senior procurement executive for the agency determines in writing that exceptional circumstances justify the use of a transaction that provides for innovative business arrangements or structures that would not be feasible or appropriate under a contract, or would provide an opportunity to expand the defense supply base in a manner that would not be practical or feasible under a contract." (NDAA, 2015).
[32] This is allowed provided competitive procedures were used in the original award and the contractor successfully completed the prototype project (NDAA, 2015)
[33] Fifteen programs were inducted into the pilot; best practices are summarized in the Agile Software Acquisition Guidebook (Cummings, 2020).

organizational and product complexity, radical innovation performance and whether innovation occurs within a firm or cross firms matter (Dean et al., 2020).

Time to market is an important factor in commercial new product development. Browning and Yassine note that contrary to most of the literature and most models, product development is commonly cyclical (Browning & Yassine, 2016). They considered different program development policies ("priority rules") for both cyclical and acyclical program and portfolios with varying degrees of resource contention, and derived a small set of priority rules for program offices to minimize average project or portfolio delay (Browning & Yassine, 2016). Evans and Johnson developed an ordinal "innovation readiness level" (IRL) scale, providing a supplier view of innovation that included business outcomes such as beta version sales and cash-positive operations, and considered other factors such as human resources, legal, and financial readiness in their overall model; they did not, however, address time to market (Evans & Johnson, 2013).

Sherman and Rhoades noted that incentives and sanctions have to be aligned to favor cycle time reduction and provide historical government and industry examples (Sherman & Rhoades, 2010). For example, urgent acquisition programs may have broad exemptions from established statutory and regulatory controls (NDAA, 2015, sec. 803). They benefit from access to additional resources and support (Lord, 2020), which are strong incentives for government entities to find new innovations. Acquisition schedules[34] between 1997 to 2015 averaged about seven years in duration (Kendall, 2016). Several factors are related to faster defense acquisitions, such as need urgency and senior leader sponsorship (Van Atta et al., 2016); using proven systems (Tate, 2016), using non-waterfall software development methods and adapting commercial technologies (A. Etemadi & Kamp, 2021), and organizational planning and execution competence (Jaifer et al., 2020). We previously showed how suppliers in defense-unique markets are not incentivized to shorten schedule duration (A. H. Etemadi & Kamp, 2021).

Conclusions and Research Hypotheses

There are several barriers to innovation within the DoD. Traditional processes that worked well when the DoD held the largest market share now must compete for performers when commercial markets provide greater economic incentives. Congress intended Middle Tier Acquisitions to deliver prototypes or fielded systems within five years of program start (NDAA, 2015). They inserted several incentives encouraging DoD use such as reduced requirements, broader authorities, and access to resources. Our research hypothesis is that *Middle Tier Acquisition programs have shorter schedule durations than other rapid acquisition approaches*.

The next section of this paper discusses the research methodology and datasets. The paper continues with a discussion of results and conclusions.

Methodology

Schedule growth is problematic given the emphasis on shorter durations. There are two types of Middle Tier Acquisitions—rapid prototyping and rapid fielding. Figure 2 provides an example program schedule plan.

[34] Often called cycle time in the literature.

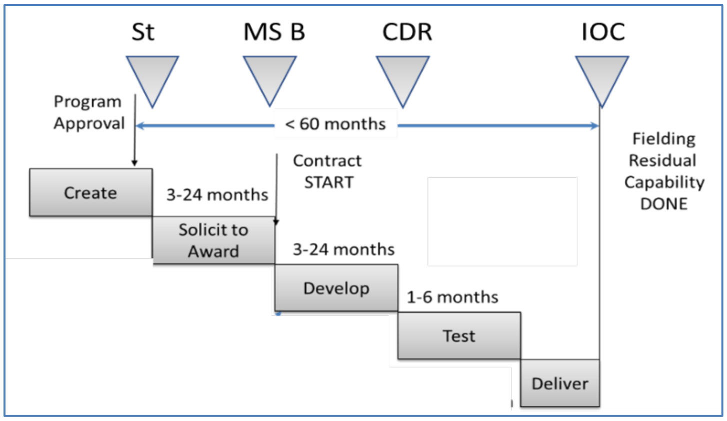


Figure 11. Middle Tier Acquisition Rapid Prototyping Schedule Model

Figure 2 shows four gates or milestones—program approval or start (St), development start (MS.B), design review (CDR), and delivery (IOC). The durations and events in Figure 2 are notional, and may be changed to meet programmatic objectives. This model has three *intervals* or *phases*—the time between approval and development start (St.B), the time from development start to design review (B.CDR), and the time from design review to delivery (CDR.IOC). Schedule growth may occur during one or more of these intervals. Table 2 shows common causes of schedule growth in each interval.

Table 3. Interval Schedule Definitions and Literature Growth Causes

Interval	Causes for Schedule Growth	Reference
Approval to development start (St.B)	Contracting issues	(Riposo et al., 2014) (Asadabadi & Sharpe, 2019)
Development start to design review (B.CDR)	Technology maturity Requirements uncertainty	(Katz et al., 2015) (Fernandes et al., 2015)
Design review to delivery (CDR.IOC)	Integration and test issues	(Manuel, 2019)

We used interval duration changes as a proxy for process innovation. This provided insight into not only overall process change but where improvements occurred.

We used publicly available data sources for this research: General Accountability Officer (GAO) annual weapon system assessments, released Selected Acquisition Reports (SARs), Director, Operational Test and Evaluation (DOTE) Annual Reports, and data from FPDS.gov and usaspending.gov websites. We created a dataset using the 2020 GAO annual weapon system assessment (n= 63; Dodaro, 2020), and eliminated entries with insufficient data or changing structures, leaving 53 entries. Table 3 summarizes the dataset.

Table 4. Selected GAO 2020 Programs

Type	AIR	C3I	GND	MSL	SHIP	SPACE
MCA	APT	AMDR	**ACV**	AARGM-ER	**CVN78**	WSF
	B2DMSM	**HMS**	AMPV	***SDB.INC2***	**DDG1000**	
	CIRCM	***OCX.BLK1.2***		**JAGM**	**FFGX**	
	CRH	***JPALS***		**IFPC.Inc2**	**SSBN826**	
	F15EPAWSS	***IAMD***		PrSM	**SSC**	
	CH-53K				TAO205	
	KC46A				**DDG51FLT3**	

	IRST.BLK2				LHA8	
	UH-1N.REP				LPD17	
	MQ25				**SSN774.BLK5**	
	MQ4C					
	NGJ-MB					
	ITEP					
	VC25.RECAP					
	VH92					
	B52RMP					
	P8A.INC3					
	F35					
MTA	**B52CERP**	**LTAMDS**	***ERCA.Inc1C***	*ARRW*		*OPIR.BLK0*
	F22CP	*UP*	***IVAS***	HCSW		*PTES*
			MPF			*PTS*
Coding	No code	*Agile*	**Modular**	***Agile+Modular***		

The program types are Major Capability Acquisition[35] (MCA) and Middle Tier Acquisition (MTA; Lord, 2020). The columns sort the programs into commodity types. The programs were coded as modular or Agile development based on review of the public reports. Contract data was substantial, programmatic data sparse. We manually validated the dataset[36]. We compared program types by coding using graphical methods. Descriptive statistics and quantitative tests were used to quantify and confirm significance for sufficient populations, and we used Mood's Median Test to test the research hypothesis. The next section presents the analysis results.

Results And Analysis

We sorted the programs by MCA or MTA and by coding as an Agile or Modular development program. We start by comparing start phase durations between different program types. Figure 3 shows the 2020 start phase (St.B) intervals for all MCAs and MTAs, and the cumulative distribution of St.B for modular and Agile MCA developments.

[35] This is also known as a Major Defense Acquisition Program (MDAP).
[36] In some cases, policy delayed public release.

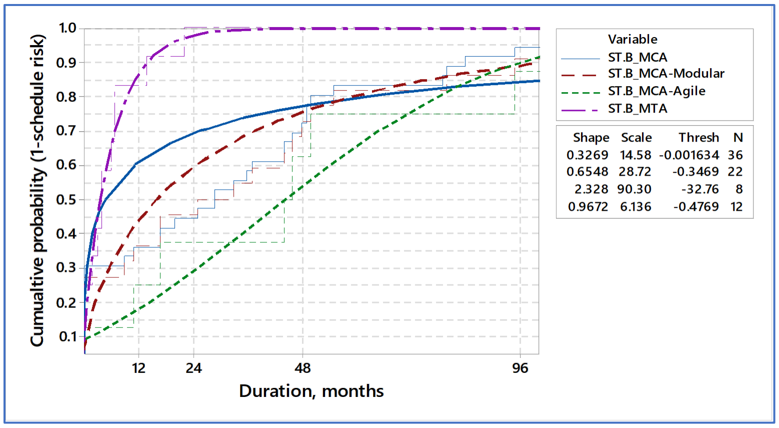


Figure 12. Start Intervals.
(GAO, 2020).

The interval between program start and development start St.B depends on the source selection and award process. Note that half of the MCA programs achieved Milestone B within 12 months, while about half of MTAs achieved Milestone B in less than six months. A Mood's median test shows MTA start intervals are statistically different (χ^2 = 10.52, p-value = 0.018) than the start interval medians of the various types of MCAs. Therefore, the MTA solicitation and awards process is different than traditional processes as MTAs are more likely than MCAs at awarding contacts in less than 12 months. Table 4 shows modular program phase statistics.

Table 5. Modular Program Interval Durations Summary.
(GAO, 2020).

		Modular					Non-modular				
Interval	Type	N	N*	Mean	StDev	Median	N	N*	Mean	StDev	Median
ST.B	MCA	22	2	35.55	35.54	29	14	3	29.93	28.92	29
	MTA	5	0	3.6	6.07	0	7	0	7.29	6.78	6
B.CDR	MCA	19	5	28.42	27.33	21	17	0	23.94	18.04	18
	MTA	2	3	16.5	10.61	16.5	3	4	22.33	13.05	18
CDR.IOC	MCA	18	6	89.78	35.98	88	17	0	67.76	20.95	59
	MTA	2	3	43.5	10.61	43.5	3	4	34	12.77	31

The small populations make graphical analysis more insightful, and quantitative analysis less meaningful. Figure 4 shows the program intervals for modularity-coded MCAs and MTAs.

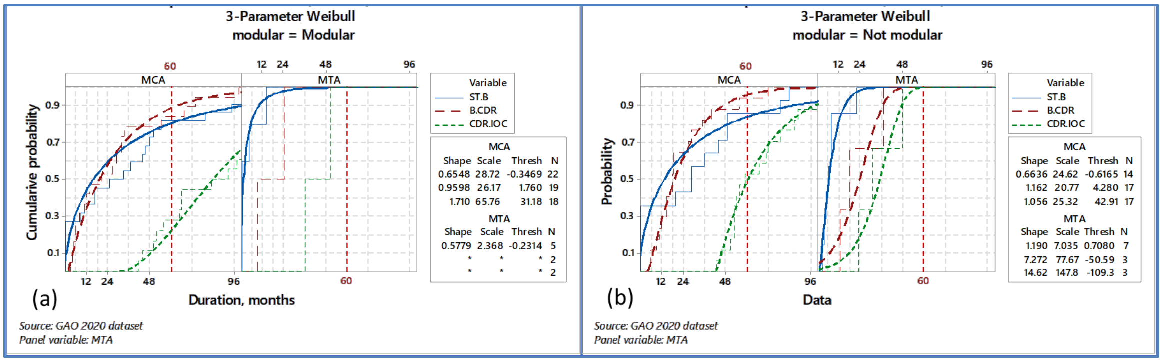


Figure 13. Modular Development Inter-Event Durations, Weibull Distribution

Figure 4a shows the modular-coded program phase results, and Figure 4b shows the non-modular coded phase results. Modular MCA CDR.IOC phases are longer than the first two phases, and that MTA start phase is fast relative to an MCA. No useful qualitative results can be drawn for B.CDR and CDR.IOC for modular-coded MTAs. For non-modular coded programs, the MTA St.B phase looks similar to the modular programs. The non-modular B.CDR and CDR.IOC phases complete sooner than for non-modular MCAs. However, non-modular development (B.CDR) MCA and MTA phases are closer in duration, with more than 90% of all programs completing B.CDR in less than 48 months. Table 5 shows Agile program phase statistics for both program types.

Table 6. Agile Program Interval Durations Summary.
(GAO, 2020).

		Agile					Non-Agile				
Interval	Type	N	N*	Mean	StDev	Median	N	N*	Mean	StDev	Median
ST.B	MCA	8	1	46.9	39.6	45	28	4	29.5	30.33	27
	MTA	9	0	3.222	2.95	3	3	0	13.33	9.02	14
B.CDR	MCA	8	1	30	27.49	22	28	4	25.25	22.28	18.5
	MTA	5	4	20	11.11	18	0	3	*	*	*
CDR.IOC	MCA	8	1	101.3	32.2	111.5	27	5	72.52	28.37	65
	MTA	5	4	37.8	11.69	36	0	3	*	*	*

Figure 5 shows Agile-coded program intervals.

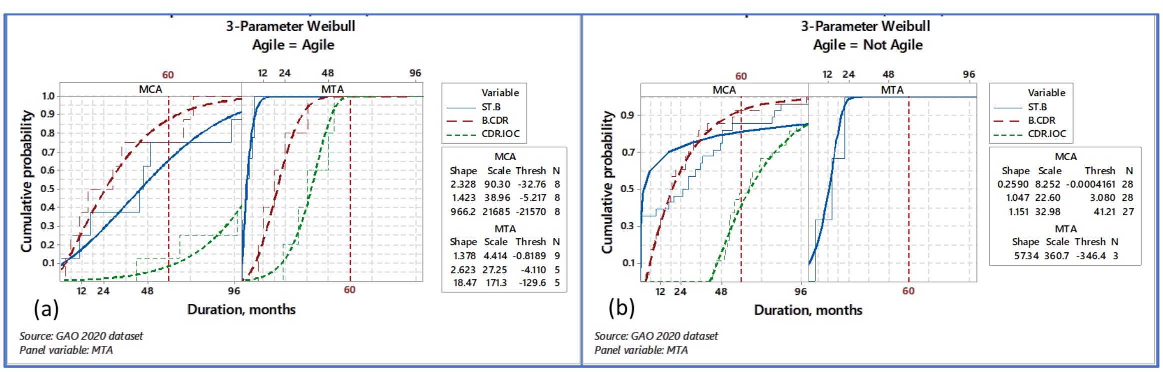


Figure 14. Agile-Inter-Event Durations, Weibull Distribution

The Agile-coded MCA start (St.B) phases were slower than their development (B.CDR) phases. Non-Agile MTAs did not have a specified design review, so they have a different program structure than the Agile-coded MTAs. These are all summarized in Table 6.

Table 7. Overall Program Interval Duration Summary. (GAO, 2020).

				Overall		
Interval	**Type**	**N**	**N***	**Mean**	**StDev**	**Median**
St.B	MCA	36	5	33.36	32.81	29
	MTA	12	0	5.75	6.48	4
B.CDR	MCA	36	5	26.31	23.19	18.5
	MTA	5	7	20	11.11	18
CDR.IOC	MCA	35	6	79.09	31.28	67
	MTA	5	7	37.8	11.69	36

The overall statistics are reasonable for comparing groups. Note that median St.B and CDR.IOC durations are smaller for MTAs than for MCAs, while B.CDR durations are similar. No Agile-coded MTA programs had identified design reviews. Figure 6 compares the different phase cumulative distributions.

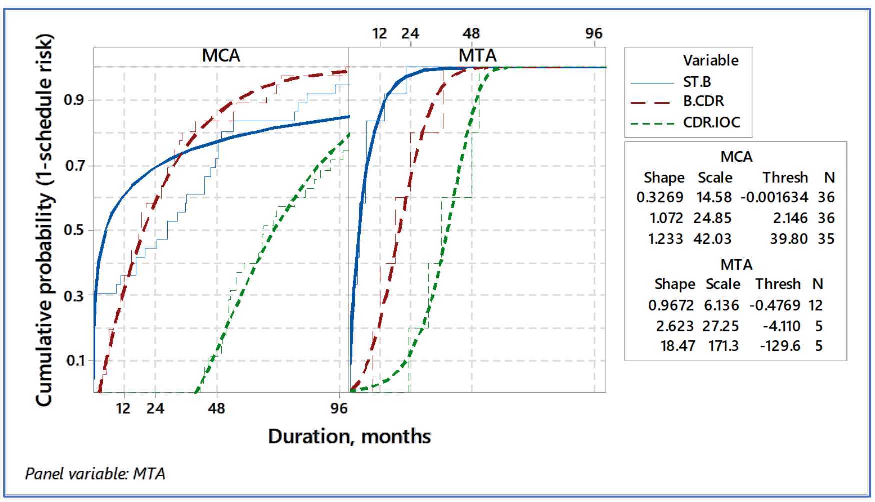


Figure 15. Overall Interval Cumulative Distributions

The figure shows the much faster start (St.B) and delivery (CDR.IOC) phases of MTAs relative to traditional programs. It also shows that the development (B.CDR) phases are similar, again with more than 90% of all programs completing design review within four years of development start. It also shows that for an MTA to achieve its objective of delivery within 60 months of start, it must have a very fast (less than three month) start (St.B) phase, a development phase of less than two years, leaving the remainder of about three years for delivery. About four in 10 (about 40%) of MTAs could achieve these goals assuming serial phases. Table 7 looks at how phase durations correlated with prior significant cycle time predictor variables (A. H. Etemadi & Kamp, 2021).

Table 8. Factor Correlations.
(GAO, 2020).

	Interval or factor	St.B	A	B	C	D	E	F	G	H
A	B.CDR	--								
B	CDR.IOC	--	**0.41							
C	Cycle time, months	***0.55	**0.42	***0.78						
D	R&D Budget	*0.27	*0.30	*0.35	***0.50					
E	% Change in (D)	*0.31	--	**0.50	***0.59	*0.34				
F	Budget Importance	--	*0.27	***0.68	***0.67	***0.53	*0.29			
G	Unit cost	--	*0.33	--	*0.32	--	--	--		
H	% change in buy	**_*-0.25_**	--	*0.29	--	**_*-0.30_**	**_**-0.41_**	--	--	
I	% change in (C)	--	--	***0.50	***0.51	--	***0.64	*0.24	--	--

*p-values *0.xxx - < 0.1 **0.xxx - < 0.01 ***0.xxx - <0.001*

Negative correlations in Table 7 are in ***bold italics***. Table 7 shows that interval durations correlated with research and development budgets, because major program overall schedule are correlated with these factors. *No correlations with prior significant MCA predictor variables were found for an MTA-only dataset*[37]. Finally, we used Mood's Median Test to test if medians were statistically different between MCA and MTA program phases. Table 8 shows the results of this testing.

Table 9. Mood's Median Test Summary

	Group	Median	N <= Overall	N > Overall	Q3 – Q1	95% Median CI	DF	X^2	P-Value
St.B	MCA	29	13	23	49	(10.5, 44.5)	1	11.11	**0.001**
	MTA	4	11	1	6.5	(0.5, 7)			
	*Overall	15.5							
B.CDR	MCA	18.5	18	18	24	(12, 29)	1	0.18	0.675
	MTA	18	3	2	20	(9, 37)			
	*Overall	18							
CDR.IOC	MCA	67	15	20	55	(56, 88.3905)	1	5.71	**0.017**
	MTA	36	5	0	22.5	(23, 51)			
	*Overall	65.5							

Two phase intervals were statistically different—the start (St.B) and the deployment (CDR.IOC) phases. The test did not find a significant difference in the development (B.CDR) phase.

[37] See Appendix for details.

Conclusions

All research questions were addressed. Middle Tier Acquisitions have significantly shorter start (St.B) and deployment (CDR.IOC) phases than Major Capability Acquisitions. Additionally, there were no correlations between Middle Tier Acquisitions phase durations and known correlates with Major Capability Acquisitions. These phase differences are due to using acquisition authorities such as commercial-like contracting methods[38], acquisition tailoring, and limited production runs to satisfy delivery definitions. While MTAs may include modular or Agile development methods or principles, the statutes incentivize limiting explicit requirements, delivered quantities, and testing activities. The technical risk of MTA programs is implicitly limited by the statutory duration limit, incentivizing program offices and contractors towards technologies and products deliverable within this limit. Commercial New Product Development technical risk constraints are similarly driven by time-to-market and budget limits, but the motivation is profit or loss instead of statutory limits.

Middle Tier Acquisition programs *have shorter schedule durations than other rapid acquisition approaches (research hypothesis).* Modular development schedules may be longer than other equivalent programs due to testing and validating the modular interfaces and interactions. Following initial delivery, subsequent changes may be less complex. Agile development moves quickly, but MTAs have an explicit transition to sustainment which makes the MTA deployment phase faster.

The Middle Tier Acquisition pathway provides structural incentives for programs to deliver capabilities in a short period of time. They complement existing rapid acquisition processes and highlight the importance of aligning incentives and objectives.

Future research should revisit the FY 2020 Middle Tier Acquisitions and confirm or refute predicted outcomes. Access to non-public information such as program strategies and surveys of program office personnel would illuminate the underlying decisions and trades made for different types of rapid acquisition programs.

References

10 U.S.C. § 2371—Research projects: Transactions other than contracts and grants, Pub. L. No. 103–160, 2371 10 (1993).

10 U.S.C. §2430—Major Defense Acquisition Program Defined, Pub. L. No. 111–84, 10 U.S.C. § 2430 U.S. Code (2021). https://www.gpo.gov/fdsys/pkg/USCODE-2006-title10/html/USCODE-2006-title10-subtitleA-partIV-chap144-sec2430.htm

Asadabadi, M. R., & Sharpe, K. (2019). The ambiguity dilemma in procurement projects. *Journal of Business & Industrial Marketing*, *34*(4), 792–800. https://doi.org/10.1108/JBIM-05-2018-0157

Behler, R. F. (2019). *Director, operational test and evaluation FY 2019 annual report*. https://www.dote.osd.mil/Publications/Annual-Reports/2019-Annual-Report/

Bonnin Roca, J., & O'Sullivan, E. (2020). The role of regulators in mitigating uncertainty within the Valley of Death. *Technovation*, 102157. https://doi.org/10.1016/j.technovation.2020.102157

Braha, D., & Bar-Yam, Y. (2007). The statistical mechanics of complex product development: Empirical and analytical results. *Management Science*, *53*(7), 1127–1145. https://doi.org/10.1287/mnsc.1060.0617

Bresler, A. (2018). Improving defense innovation programs to enhance force readiness. *Journal of Defense Analytics and Logistics*, *2*(2), 110–124. https://doi.org/10.1108/JDAL-06-2018-0010

Browning, T. R., & Yassine, A. A. (2016). Managing a portfolio of product development projects under resource constraints. *Decision Sciences*, *47*(2), 333–372. https://doi.org/10.1111/deci.12172

[38] Such as 10 USC 2371b (Other Transaction Agreements) and 10 USC 2373 (Procurements for Experimental Purposes).

Caiazza, R. (2016). A cross-national analysis of policies affecting innovation diffusion. *The Journal of Technology Transfer*, *41*(6), 1406–1419. https://doi.org/10.1007/s10961-015-9439-2
Cha, C. (2016). *DoD major automated information systems improvements can be made in reporting critical changes and clarifying leadership responsibility* (GAO-16-336; p. 57). Government Accountability Office. https://www.gao.gov/assets/680/676189.pdf
Cummings, S. (2020). *Agile software acquisition guidebook*. Department of Defense. https://www.dau.edu/cop/it/_layouts/15/WopiFrame.aspx?sourcedoc=/cop/it/DAU%20Sponsored%20Documents/AgilePilotsGuidebook%20V1.0%2027Feb20.pdf&action=default
de Rassenfosse, G., Jaffe, A., & Raiteri, E. (2019). The procurement of innovation by the U.S. government. *PloS One*, *14*(8), e0218927–e0218927. https://doi.org/10.1371/journal.pone.0218927
Dean, T., Zhang, H., & Xiao, Y. (2020). The role of complexity in the Valley of Death and radical innovation performance. *Technovation*, 102160. https://doi.org/10.1016/j.technovation.2020.102160
Defense Business Board. (2015). *Transforming DoDs core business processes for revolutionary change* (No. 15–01; p. 148). Department of Defense. https://dbb.defense.gov/Portals/35/Documents/Reports/2015/DBB%20FY15-01%20Core%20Business%20Processes.pdf
Defense Innovation Unit. (2020). https://diu.mil
Dodaro, G. L. (2020). *Defense acquisitions annual assessment drive to deliver capabilities faster increases importance of program knowledge and consistent data for oversight* (GAO-20-439; p. 251). Government Accountability Office. https://www.gao.gov/products/GAO-20-439
Drabkin, D., Ahern, D., Branch, E., Dyer, J., Hoskin, M., Merchant, K. D., Trowel, L., Raney, T. L., Thompson, N. R., Blake, C. D., Burman, A. V., Garman, C. D., LaPlante, W. A., Metzger, D. P., & Scott, D. A. (2016). *Section 809 panel*. Defense Technical Innovation Center. https://discover.dtic.mil/section-809-panel/
Etemadi, A. H., & Kamp, J. (2021). Market and contractor factors affecting rapid acquisition strategies. *Systems Engineering*. https://doi.org/10.1002/sys.21577
Etemadi, A., & Kamp, J. (2021). Acquisition strategy factors related to faster defense acquisitions. *Systems Engineering*. https://doi.org/10.1002/(ISSN)1520-6858
Evans, J. D., & Johnson, R. O. (2013). Tools for managing early-stage business model innovation. *Research Technology Management*, *56*(5), 52–56. https://doi.org/10.5437/08956308X5605007
Fagerberg, J. (2017). Innovation policy: Rationales, lessons, and challenges. *Journal of Economic Surveys*, *31*(2), 497–512. https://doi.org/10.1111/joes.12164
Fernandes, J., Henriques, E., Silva, A., & Moss, M. A. (2015). Requirements change in complex technical systems: An empirical study of root causes. *Research in Engineering Design*, *26*(1), 37–55. https://doi.org/10.1007/s00163-014-0183-7
Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, *30*(7), 1019–1039. https://doi.org/10.1016/S0048-7333(00)00135-9
Flournoy, M. A., & Lyons, R. R., III. (2016). Sustaining and enhancing the US military's technology edge. *Strategic Studies Quarterly : SSQ*, *10*(2), 3–13.
Freedburg, S. J. (2019, December 13). Soldiers, coders surprise Army brass by changing IVAS goggles. *Breaking Defense*. https://breakingdefense.com/2019/12/soldiers-coders-surprise-army-brass-changing-ivas-goggles/
General Services Administration. (2019). *Federal acquisition regulations*. http://farsite.hill.af.mil/reghtml/regs/far2afmcfars/fardfars/far/Far1toc.htm#TopOfPage
General Services Administration. (2020, March 17). *Product and service code manual*. Acquisition.Gov. https://www.acquisition.gov/psc-manual
Hills-Duty, R. (2018, May 1). Microsoft HoloLens sales figures revealed. *VR Focus*. https://www.vrfocus.com/2018/05/microsoft-hololens-sales-figures-revealed/
Katz, D. R., Sarkani, S., Mazzuchi, T., & Conrow, E. H. (2015). The relationship of technology and design maturity to DoD weapon system cost change and schedule change during engineering and manufacturing development. *Systems Engineering*, *18*(1), 1–15. https://doi.org/10.1111/sys.21281
Kendall, F. (2014). *Better buying power 3.0* (p. 10) [White paper]. Department of Defense. http://www.acqnotes.com/Attachments/Better-Buying-Power-3.0-White-Power.pdf
Kendall, F. (2016). *Performance of the defense acquisition system 2016 annual report* (p. 224). Department of Defense.
Lockhart, D. (2018). *Open innovation: Accelerating innovative products and services through the Department of Defense acquisition management system* [Univeristy of Maryland University Campus]. https://search.proquest.com/docview/2095885707/?pq-origsite=primo

Lopez, T. (2021, June 30). Risk aversion impedes hypersonics development. *Defense.Gov*. https://www.defense.gov/News/News-Stories/Article/Article/2243224/risk-aversion-impedes-hypersonics-development/

Lord, E. M. (2020). *Operation of the adaptive acquisition framework*. Department of Defense. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500002p.pdf?ver=2019-10-25-134150-283

Lunden, I. (2014, March 27). Microsoft paid up to $150M to buy wearable computing IP from The Osterhout Design Group. *Tech Crunch*.

Manuel, D. (2019). *Integrate or innovate: Controlling schedule risk in major defense acquisition programs* [The George Washington University]. https://search.proquest.com/docview/2159035140/?pq-origsite=primo

Markham, S. K., Ward, S. J., Aiman-Smith, L., & Kingon, A. I. (2010). The valley of death as context for role theory in product innovation. *The Journal of Product Innovation Management*, *27*(3), 402–417. https://doi.org/10.1111/j.1540-5885.2010.00724.x

Meeker, M. (2018). *Internet trends 2018* (p. 294). Kleiner Perkins. https://www.kleinerperkins.com/perspectives/internet-trends-report-2018/

Mounier, E. (2011). *Technical and market trends for microbolometers for thermography and night vision*. *8012*(1), 80121U-80121U – 6. https://doi.org/10.1117/12.877050

National Defense Authorization Act of Fiscal Year 2011, Pub. L. No. 111–383, 2359a 10 (2011). https://www.law.cornell.edu/uscode/text/10/2359a

National Defense Authorization Act for Fiscal Year 2016, Pub. L. No. 114–92 (2015), S.1356. https://www.congress.gov/114/plaws/publ92/PLAW-114publ92.pdf

National Defense Authorization Act for Fiscal Year 2018, Pub. L. No. 115–91, 740 (2017). https://www.congress.gov/115/plaws/publ91/PLAW-115publ91.pdf

OECD. (2019). *Measurement of scientific and technological activities, Oslo manual: Guidelines for collecting and interpreting innovation data.* OECD Publishing. https://www-oecd-ilibrary-org.proxygw.wrlc.org/sites/9789264304604-en/1/1/4/index.html?itemId=/content/publication/9789264304604-en&_csp_=f0a6f52d4530c0667c4c56b36905227f&itemIGO=oecd&itemContentType=book#execsumm-d1e476

OUSD(R&E). (2020, August). *Defense innovation marketplace connecting industry and the Department of Defense*. Defense Innovation Marketplace. https://defenseinnovationmarketplace.dtic.mil/business-opportunities/rapid-innovation-fund/

Procurement for experimental purposes, 2373 10 (2015).

Riposo, J., McKernan, M., & Duran, C. K. (2014). *Prolonged cycle times and schedule growth in defense acquisition: A literature review*. RAND.

Roberts, J. (2014). Gigaom: Microsoft bucks its own history, embraces “openness” in push for Azure adoption. *Newstex Trade & Industry Blogs*. https://gigaom.com/2014/06/18/microsoft-bucks-its-own-history-embraces-openness-in-push-for-azure-adoption/

Schoeni, D. E. (2018). Whither innovation? Why open systems architecture may deliver on the false promise of public-private partnerships. *Administrative Law Review*, *70*(2), 409.

Sherman, J. D., & Rhoades, R. G. (2010). Cycle time reduction in defense acquisition. *Research-Technology Management*, *53*(5), 46–54. https://doi.org/10.1080/08956308.2010.11657650

Taylor, A. J. (2019). The revolution in federal procurement, 1980–present. *Business and Politics*, *21*(1), 27–52. https://doi.org/10.1017/bap.2018.9

Turkina, E., Oreshkin, B., & Kali, R. (2019). Regional innovation clusters and firm innovation performance: An interactionist approach. *Regional Studies*, *53*(8), 1193–1206. https://doi.org/10.1080/00343404.2019.1566697

Appendix

Table 10. Dataset Summary (N=53).
(GAO, 2020).

		St.B			B.CDR			CDR.IOC			Cycle.Mo		
MCA (MDAP)	term	Agile	Not Agile	All	Agile	Not Agile	All	Agile	Not Agile	All	Agile	Not Agile	All
Modular	μ (mo)	47	30	36	32	26	28	100	83	90	144	135	138
	σ (mo)	42.7	31.8	35.5	29.0	27.4	27.3	34.6	36.8	36.0	41.2	65.3	57.6
	n	8	16	24	8	16	24	8	16	24	8	16	24
Not modular	μ (mo)	44	29	30	15	25	24	108	65	68	123	92	94
	σ (mo)	*	29.8	28.9	*	18.5	18.0	*	18.8	21.0	*	33.3	33.1
	n	1	16	17	1	16	17	1	16	17	1	16	17
All	μ (mo)	47	30	33	30	25	26	101	73	79	142	113	120
	σ (mo)	39.6	30.3	32.8	27.5	22.3	23.2	32.2	28.4	31.3	39.2	55.6	53.3
	n	9	32	41	9	32	41	9	32	41	9	32	41
MTA		Agile	Not Agile	All	Agile	Not Agile	All	Agile	Not Agile	All	Agile	Not Agile	All
Modular	μ (mo)	0	9	4	17	*	17	44	*	44	49	39	45
	σ (mo)	0.0	7.1	6.1	10.6	*	10.6	10.6	*	10.6	10.2	27.6	16.7
	n	3	2	5	3	2	5	3	2	5	3	2	5
Not modular	μ (mo)	5	22	7	22	*	22	34	*	34	52	10	46
	σ (mo)	2.1	*	6.8	13.1	*	13.1	12.8	*	12.8	16.4	*	21.8
	n	6	1	7	6	1	7	6	1	7	6	1	7
All	μ (mo)	3.222	13.333	5.75	20	*	20	37.8	*	37.8	51.11	29	45.58
	σ (mo)	2.949	9.018	6.482	11.11	*	11.11	11.69	*	11.69	14	25.51	19
	n	9	3	12	9	3	12	9	3	12	9	3	12

Table 11. Interval and Factor Pearson Correlations (n=53).
(GAO, 2020).

		ST.B	B.CDR	CDR.IOC	Cycle.Mo	LN.RD.M	RD.M.PCT	BUDGET.IMP	LN.UC.M	P_no.PCT
B.CDR	All	--								
	MCA	--								
	MTA	--								
CDR.IOC	All	--	*0.41							
	MCA	--	*0.434							
	MTA	--	*-0.908							
Cycle.Mo	All	***0.548	**0.415	***0.777						
	MCA	**0.451	**0.415	***0.741						
	MTA	--	--	--						
LN.RD.M	All	*0.271	*0.298	*0.349	***0.503					
	MCA	--	*0.284	*0.403	***0.559					
	MTA	--	--	--	--					
RD.M.PCT	All	*0.314	--	**0.5	***0.591	*0.337				
	MCA	--	--	**0.498	***0.614	*0.335				
	MTA	+	+	+	+	+				
BUDGET.IMP	All	--	*0.269	***0.683	***0.666	***0.527	*0.293			
	MCA	--	--	***0.596	**0.479	***0.657	--			
	MTA	+	+	+	+	+	+			
LN.UC.M	All	--	*0.332	--	*0.321	--	--	--		
	MCA	--	*0.341	*0.321	**0.43	--	--	--		
	MTA	--	--	--	--	--	+	+		
P_no.PCT	All	*-0.246	--	*0.285	--	*-0.296	*-0.412	--	--	
	MCA	*-0.283	--	--	--	*-0.324	**-0.442	--	--	
	MTA	--	+	+	--	--	+	+	--	
Cy.Mo.PCT	All	--	--	**0.495	***0.505	--	***0.64	**0.237	--	--
	MCA	--	--	**0.451	**0.424	--	***0.627	--	--	--
	MTA	+	+	+	+	+	+	+	+	+
	*p-values *0.xxx - < 0.1 **0.xxx - < 0.01 ***0.xxx - + - NA*									

Assessing Policy Changes on the Cost of Husbanding Services for Navy Ships

Dr. Margaret Hauser—is a faculty associate at the Naval Postgraduate School. [margaret.hauser@nps.edu]

Dr. Geraldo Ferrer—is a professor at the Naval Postgraduate School. [gferrer@nps.edu]

Dr. Robert Mortlock—is a professor at the Naval Postgraduate School. [rfmortlo@nps.edu]

Abstract

Abstract: In the wake of a major corruption conspiracy, the U.S. Navy reformed husbanding services procedures to increase competition, auditability, and accountability with the end goal of reducing expenditures. The first policy change, Off-Ship Bill Pay (OSBP), formalized a process for procuring, rendering, and paying for husbanding services to increase oversight. The second policy change increased the use of multiple award contracts (MACs) in which multiple vendors are awarded a contract over a region, increasing competition for individual port visits. The purpose of this paper is to analyze the effects of these policy changes on the cost of husbanding services. Multiple regression was used to account for port visit characteristics that affect cost such as ship type and the number of days in port. MACs demonstrated a reduction effect on the cost of port visits. Further, OSBP appears to have a negligible effect on port visit cost after the initial learning curve for both Navy personnel and vendors.

Executive Summary

U.S. Navy vessels routinely make visits in non-U.S. ports for numerous reasons including resupply efforts, multi-national exercises, and liberty where husbanding services must be contracted from a commercial vendor, referred to as a husbanding service provider (HSP). Husbanding services include tugboats to guide vessels into and out of port, transportation services, waste removal and disposal, fuel, food and water, and force protection equipment and services. In the wake of the "Fat Leonard Scandal" with Glenn Defense Marine Asia (GDMA)—a major corruption conspiracy, the U.S. Navy reformed husbanding services procedures to increase competition, auditability, and accountability with the end goal of reducing expenditures. The first policy change, Off-Ship Bill Pay (OSBP), formalized a process for procuring, rendering, and paying for husbanding services to increase oversight. The second policy change increased the use of multiple award contracts (MACs) in which multiple vendors are awarded a contract over a region, increasing competition for individual port visits. A Global MAC was awarded by NAVSUP in fiscal year (FY) 2021 but does not include data from this time period. The purpose of this paper is to analyze the effects of these policy changes on the cost of husbanding services.

Historical HSP data from the HSPortal (formerly LogSSR) corresponding to port visits made starting on October 1, 2009 (FY 2010) to the last port visit arriving on June 11, 2020 was processed to filter out non-normal (e.g., maintenance, transit, brief stop for fuel, etc.) or cancelled visits. Multiple regression was used to account for port visit characteristics that affect cost such as ship type and the number of days in port. The response variable is the natural log of total cost meaning that *the regression models provide a base value for the total cost of a port visit* and the explanatory variables are multipliers to the base cost. The explanatory variables that showed a significant effect on the total cost are exhibit line item number (ELIN) count, type of mooring (pier side or anchorage), ship type, days in port, time (FY as a categorical variable), and contract type (single award contract [SAC], single visit contract [SVC], or MAC). To supplement the port visit data, the historical crude oil prices (nominal) are also included based on the date the port visit was planned. Two regression analyses are performed on the dataset assuming fixed effects. The first analysis (referred to as the “global cost model”) evaluates the entire dataset in a single model using two-stage weighted least squares regression. The global cost model assumes a fixed factor effect over the time horizon. To test this assumption, the second analysis explores a unique two-stage weighted least squares regression model for each FY; these cost models are referred to as “FY cost models.” The data is an unbalanced panel meaning that not every variable instance occurs in each FY.

The global cost model shows general trends in the total cost of port visits made FY 2010 to FY 2020. After Leonard Francis’s arrest in September 2013, the cost of husbanding services continues to increase to a peak in FY 2016. Leonard Francis’s company, GDMA, was the HSP for more than 25% of the port visits made prior to FY 2014. This is influential because after Francis’s arrest, the U.S. Government no longer did business with GDMA removing the HSP that provided services for a quarter of their port visits from the pool of possible vendors, reducing competition. Additionally, it was known that GDMA had a monopoly on services in certain ports, which made those ports no longer accessible to Navy vessels. The prohibition of business with GDMA and the restriction on ports likely contributed to the increase in total port visit cost beginning in FY 2014 to FY 2016.

The use of a MAC reduces the total cost by 16.9% while the use of an SVC increases the cost by 46.4% relative to port visit made under a SAC. The use of anchorage mooring leads to a cost increase of more than 30% compared to pier side mooring. Each of these factors had a statistically significant effect. Spending two days in port adjusts the base cost to approximately 60% of its original value and the effect increases total cost 115% by spending 10 days in port as opposed to five. There is a substantial increase in the total cost for high ELIN counts. For example, an ELIN count of 70 approximately triples the cost of the port visit; an ELIN count of 100 increases the port visit total cost nearly 570%. The higher the ELIN count, the more likely it is that costly ELINs are included. Although the effect of the price of oil is statistically significant, it does not have a large magnitude like the days in port and ELIN count effects.

The multiplier value for ship type (from the base of a DDG) appears to be correlated to tonnage for most ship types.

The purpose of the FY cost models is to test the assumption that explanatory variables (such as mooring, days in port, ship type, etc.) have constant effects over time. The factors of days in port, ELIN count, and price of oil have relatively constant effects over time. The effect of anchorage mooring appears to have a dynamic effect over time. Additionally, the MAC and SVC effects are also dynamic over time. The decreasing cost reduction power of the MAC may be due to HSPs with SAC trying to provide more competitive prices to maintain their contracts. The FY cost model demonstrates that total costs by ship type do not follow the general trend of port visit costs over time. The changing fluctuation in the effect of ship types in combination with other findings would seem to indicate that there are widely varying costs to port visits. Further, the variations cannot be completely described by using the explanatory variables in the models (days in port, ELIN count, price of oil, FY, mooring type, contract type, ship type, and port). Depending on the purpose of the analysis, useful insights can be derived from each cost model analysis. The global cost model has the advantage of aggregating the entire dataset to provide an average effect over time. Although mooring, SVC, and MAC show dynamic effects over time, the global cost model shows the aggregated effect over time of each factor. The mooring effect in the FY cost models is by far the most curious with the massive increase in FY 2015 and relatively mild effects all other FYs.

The OSBP process has been criticized for the increased administrative requirements for port visits both for Navy personnel and HSPs. OSBP may have increased port visit costs initially due to the learning curve but in recent years has not demonstrated such an effect.

The cost of husbanding services has decreased since FY 2016, coinciding with increased utilization of MACs. The decreasing cost may not be directly caused by MACs (particularly in FY 2018); however, there is likely an indirect effect from MAC contracts that has motivated HSPs operating under SACs with the Navy to provide more competitive prices in order to maintain the SAC in their designated markets. This is a promising indication since NAVSUP awarded a Global MAC on October 2, 2020.

This study has a few limitations. The variation in the total cost of port visits is not completely captured by either model presented in this study; there are characteristics or components of port visits that are unknown and unaccounted for in the models. In addition, it is likely there are still port visits that were either cancelled or are non-normal port visits that remain in the dataset due to lack of identification; these observations pollute the model. Finally, the selection of model type depends on the desired perspective of the data. For an overall impression of the effect of certain factors, such as contract type or the number of days in port, the global cost model provides the best estimate since it is aggregated over the entire dataset. From this claim, the authors have drawn the conclusions of a decreasing effect on cost achieved by the OSBP as well as a cost reducing effect from the MAC. The FY cost models have reduced statistical significance for certain variables due to the lack or small sample of observations. Though the FY cost models provide less complete information, the FY cost models exposes a dynamic effect from certain variables, which provides insight into cost trends.

Federal Contracting: Senior Leaders Should Use Leading Companies' Key Practices to Improve Performance

Timothy J. DiNapoli—Managing Director, Contracting and National Security Acquisitions, U.S. Government Accountability Office. [dinapolit@gao.gov]

Nathan Tranquilli—Assistant Director, Contracting and National Security Acquisitions, U.S. Government Accountability Office. [tranquillin@gao.gov]

Guisseli Reyes-Turnell—Assistant Director, Contracting and National Security Acquisitions, U.S. Government Accountability Office. [reyesturnellg@gao.gov]

Abstract

Each year, federal agencies spend over $500 billion to buy a wide variety of products and services, ranging from cutting-edge military aircraft to common office supplies. Given the federal funds spent and the missions these contracts support, it is critical that agencies' procurement leaders manage their organizations effectively. GAO found procurement leaders at six of the federal government's largest agencies did not consistently use key practices that leading companies use to improve the performance of their procurement organizations. For example, only the procurement leaders at NASA collaborated with end users when developing performance metrics. Corporate procurement leaders told GAO that collaboration with end users during the development and implementation of performance metrics increases coordination and improves performance at the strategic level. Additionally, GAO found procurement leaders at most of the agencies reviewed had ongoing or planned efforts to use performance metrics to measure at least one of the four procurement outcomes identified as important by corporate procurement leaders: (1) cost savings/avoidance, (2) timeliness of deliveries, (3) quality of deliverables, and (4) end-user satisfaction. However, all of the leaders had work to do to fully implement metrics measuring these outcomes. The original GAO report is accessible at www.gao.gov/products/gao-21-491.

Methodology

GAO's report examined key practices that leading companies use to improve the performance of their procurement organizations, and the extent to which procurement leaders at selected federal agencies use those practices. GAO interviewed senior procurement leaders at seven leading companies, and experts from four professional associations and five academic institutions. GAO selected these individuals based on literature reviews and conversations with knowledgeable officials. GAO compared key practices they identified to those used at six federal agencies selected based on the dollar value and number of procurement actions, among other factors:

- The Air Force
- The Army
- The Navy
- The Department of Homeland Security (DHS)
- NASA
- The Department of Veterans Affairs (VA)

GAO analyzed documentation on each agency's procurement management practices, and interviewed the agencies' senior procurement leaders. The federal government does not have generally accepted definitions for outcome-oriented and process-oriented metrics.

Background

The Federal Acquisition Regulation (FAR)—which provides policies and procedures for federal government acquisition—states that the federal acquisition system must be responsive and adaptive to customer needs, concerns, and feedback (FAR 1.102-2). There are a variety of officials responsible for managing the contracting functions at federal agencies. The FAR establishes that, unless specifically prohibited by another provision of law, authority and responsibility to contract for products and services is vested in the agency head (FAR 1.601). The FAR also states that the agency head may establish contracting activities, and delegate broad authority to manage the agency's contracting functions to the Heads of the Contracting Activities (HCA). Further, the Services Acquisition Reform Act established that non-DOD agencies' Senior Procurement Executives (SPE) are generally responsible for (1) ensuring that procurement goals align with agencies' missions, (2) establishing procurement policies, and (3) managing the agencies' procurement activities (Services Acquisition Reform Act of 2003). Table 1 presents the senior procurement leaders—HCAs and SPEs—we focused on in this review.

Table 1: Senior Procurement Leaders at the Federal Agencies Included in GAO's Review

Agencies	Senior Procurement Leaders
Department of the Air Force	Deputy Assistant Secretary of the Air Force (Contracting)/Head of ContractingActivity (HCA)
Department of the Army	Deputy Assistant Secretary of the Army (Procurement)/HCA
Department of the Navy	Deputy Assistant Secretary of the Navy (Procurement)/HCA
Department of Homeland Security	Chief Procurement Officer/Senior Procurement Executive (SPE)
National Aeronautics and Space Administration	Assistant Administrator, Office of Procurement/SPE
Department of Veterans Affairs	Executive Director, Office of Acquisition and Logistics/SPE

Source: GAO analysis of agency documentation.

Note: The Army and the Navy have a total of four and 11 HCAs, respectively. In addition to the senior procurement leaders identified in this table who have department-wide responsibilities, the Army and Navy have HCAs with narrower areas of responsibility, for example, at an individual command.

Prior GAO Work on Performance Management

Congress has taken actions to improve performance management across the federal government, including the management of agencies' procurement operations. In 1993, Congress passed, and the president signed into law, the Government Performance and Results Act (GPRA) to improve agencies' performance by establishing a framework for developing and integrating agencies' missions, strategic priorities, and performance goals, among other things (GPRA, 1993). Congress subsequently amended GPRA with the GPRA Modernization Act of 2010, which includes several provisions that provide an opportunity for agencies to increase federal agencies' use of information to improve their performance (GPRA Modernization Act, 2010). We previously reported on how agencies can better meet the intent of GPRA and the GPRA Modernization Act. For example, we identified key practices agencies can take to implement these laws, including the following:

- **Linkage between individual performance and organization success:** We found that explicit linkage helps individuals see the connection between their daily activities and organizational goals, and encourages individuals to focus on how they can help achieve those goals (GAO, 2003, 2017).
- **Collaborating with stakeholders on performance management:** We found it is valuable for performance evaluators to develop relationships with stakeholders to gain

their input and buy-in, and that doing so can increase the usefulness and use of performance information in program management and policy (GAO, 1996, 2005, 2013).

- **Using performance information:** We found that agencies should establish and use a balanced set of performance measures, including outcome and process measures, and that they should obtain complete and reliable performance information (GAO, 1996, 2005, 2015). We found these actions help federal agencies identify improvement opportunities, set priorities, and allocate resources.

For the purposes of the GAO report, we established two categories of performance metrics for contracting: (1) outcome-oriented performance metrics, and (2) process-oriented performance metrics. The federal government does not have generally accepted definitions for these categories, so we defined them as follows. Outcome-oriented performance metrics for procurement organizations are those metrics that measure the results of organizations' procurement activities. Process-oriented performance metrics for procurement organizations are those metrics that measure the type or level of procurement activities conducted. Both types of measures have merit. See Table 2 for examples of outcome and process-oriented performance metrics for procurement organizations, whether in the government or the private sector.

Table 2: Examples of Performance Metrics for Procurement Organizations

Outcome-oriented	**Process-oriented**
Quality of product or services procured	Number of contract awards
Timeliness of deliveries to end users	Competition rates[a]
Cost savings or avoidance	Small business utilization rates

Source: GAO analysis of leading practices for private sector companies and agencies in our review.

[a]In general, "competition rate" measures the extent to which contracts are competitively awarded pursuant to the Federal Acquisition Regulation.

Additionally, for the purposes of this report, we are using the term *end users* to identify internal stakeholders that use the products and services procured. Some individuals identify end users as "customers," but we are not using the term *customer*, except when agencies use it in formal documentation. We chose the term end user because some key leaders told us they did not agree with the unequal partnership the term customer implies.

Differences between Procurement Organizations at Companies and Agencies

Leading companies and federal agencies both buy a wide variety of products and services critical to their operations. However, procurement leaders at leading companies operate in a different environment than procurement leaders at federal agencies. Procurement leaders at leading companies often focus on financial measures like profit margins and return on investment, but procurement leaders at federal agencies do not. Further, procurement leaders' actions are subject to laws and regulations intended to promote transparency and fairness, and to support socioeconomic goals. For example, procurement leaders are expected to maximize competition for government contracts, and meet small business utilization goals, which can introduce additional dimensions to procurement leaders' management responsibilities at federal agencies (FAR 7.105).[39]

Our prior work identified leading practices for federal agencies to obtain best value for taxpayer dollars and meet socioeconomic goals. For example, to leverage the government's buying power, we recommended that agencies enhance various category management efforts intended to improve how the government buys common products and services (GAO, 2004, 2016, 2020a). To meet socioeconomic goals, we recommended the Small Business Administration improve oversight of the women-owned small business program, and improve record-keeping of federal contracting and subcontracting opportunities for small businesses (GAO, 2019, 2020b).

Procurement Leaders at Leading Companies Generally Use Three Key Practices to Improve Organizational Performance

Corporate procurement leaders and subject matter experts we interviewed told us leading companies have increasingly recognized the extent to which procurement operations help them achieve their overarching business goals. Based on those interviews, we identified three key practices leading companies use to improve the performance of their procurement organizations and help their companies achieve strategic goals (see Figure 1).

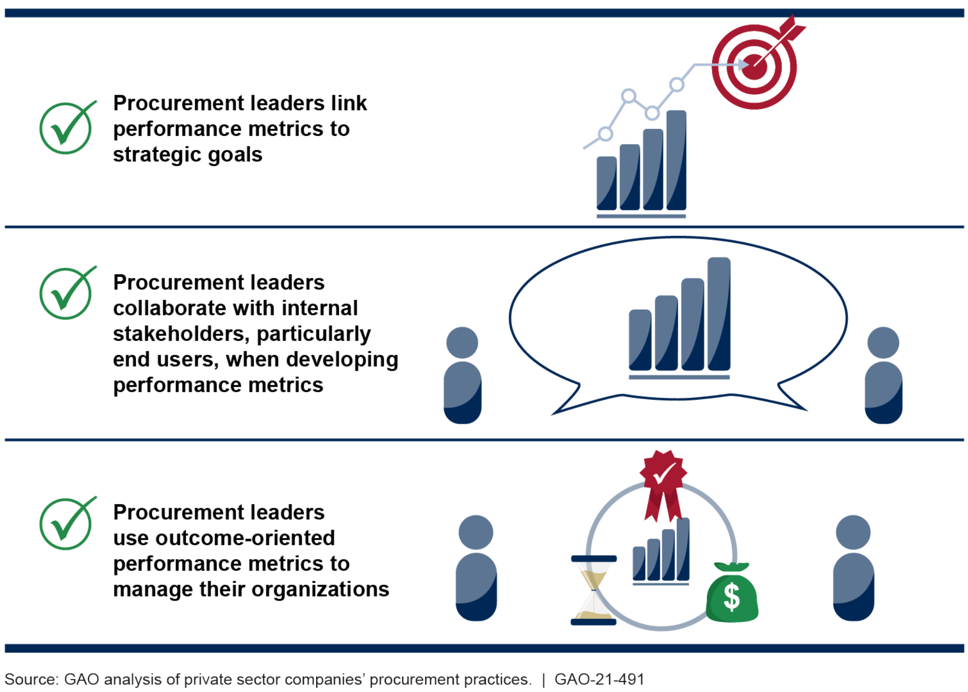


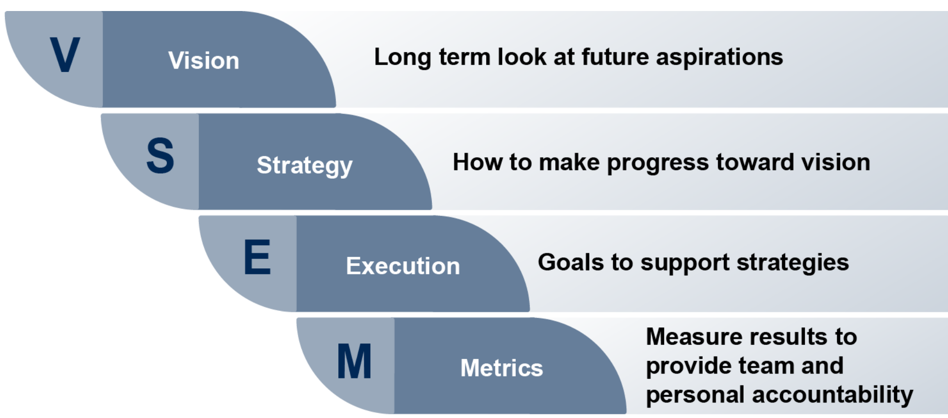
Figure 1: Procurement Leaders at Leading Companies Generally Use Three Key Practices When Managing Their Procurement Organizations

[39]Federal program managers are required to engage in acquisition planning to ensure maximum competition, while considering small business contract goals (FAR 7.105).

Corporate Procurement Leaders Link Performance Metrics to Strategic Goals

In interviews, corporate procurement leaders and subject matter experts emphasized the importance of linking performance metrics and procurement goals to corporate strategic goals. For example, a procurement expert at the Naval Postgraduate School told us the private sector has come to appreciate the extent to which procurement operations help companies achieve their overarching business goals, and that it is now common for corporate-level goals to drive procurement-specific goals and metrics. The expert added that companies can make better purchasing decisions when their procurement teams understand how they are expected to contribute to corporate goals. Similarly, a procurement executive who served as a senior contracting leader at a private technology firm, and prior to that in similar positions at several federal agencies, told us that successful private sector contracting leaders and organizations link their procurement teams' goals to the overall organization's goals. Procurement leaders from leading companies provided us examples of how they linked performance metrics to strategic goals, including the following.

- **Facebook** procurement leaders told us the company uses the Vision, Strategy, Execution, and Metrics (VSEM) method—which was originally pioneered by Cisco—to link the procurement team's metrics to the company's top-level goals. Facebook representatives said the VSEM method allows the procurement team to understand how its activities contribute to the company's overarching strategy. For example, procurement representatives told us they used the VSEM method to translate Facebook's strategic goals—which focus on quality, speed, protecting Facebook, and cost—into performance metrics (see Figure 2).



Source: GAO depiction of VSEM concept by Stewart McCutcheon and Ecolab. | GAO-21-491

Figure 2: The Vision, Strategy, Execution, and Metrics Method

- **General Electric (Aviation)** procurement leaders told us their company's leadership uses the Hoshin Kanri process to link procurement goals and performance metrics to strategic goals (see Figure 3). This process is well established, and used by other leading companies—including Toyota and Hewlett Packard—to communicate strategic goals throughout the company and link them to lower-order goals and metrics, including goals and metrics for procurement teams. For example, a strategic goal to improve product quality could drive a procurement goal for reducing defects in components procured from key suppliers, and a corresponding metric that measures the number of defects per thousand units procured.

STEP 1	Establish organizational vision
STEP 2	Develop strategic objectives
STEP 3	Develop annual objectives
STEP 4	Deploy annual objectives
STEP 5	Implement annual objectives
STEP 6	Conduct monthly review
STEP 7	Conduct annual review

Source: GAO depiction of the Hoshin Kanri process from "*The Seven Steps of Hoshin Planning*" (Waldo, Lean Methods Group). | GAO-21-491

Figure 3: The Hoshin Kanri Process

- **Procter and Gamble's (P&G)** senior procurement leader told us he ensured that there was linkage between his company's strategic goals and his procurement team's performance metrics. Specifically, he told us he uses language from the strategic goals when reviewing his procurement team's work plans, and that this approach facilitates consistent messaging, which is critical to building a team and common goals.

Procurement Leaders Collaborate with Internal Stakeholders, Particularly End Users, When Developing Performance Metrics

Corporate procurement leaders told us they also collaborate with internal stakeholders to determine what procurement performance metrics should measure. These leaders said it is particularly important to collaborate with the internal stakeholders that use the products and services their teams procure—these stakeholders are often referred to as "end users." The procurement leaders told us that collaboration with end users and end-user representatives increases coordination across functional teams—for example, sales, logistics, finance—and improves performance at the strategic level. For example:

- **Raytheon Technologies** procurement leaders told us they continually collaborate with other functional teams when establishing performance metrics and goals to ensure they do not conflict with one another. For example, a procurement team with a unit-price metric may be incentivized to buy large volumes of a commodity to get a discount rate, but this approach could conflict with a logistics team's efforts to decrease warehousing costs. Raytheon Technologies' procurement leaders told us they mitigate these types of conflicts through cross-functional coordination focused on strategic goals, such as reducing total operating costs.
- One of **ExxonMobil's** senior procurement leaders told us that procurement teams are expected to collaborate and maximize efficiencies across functional teams when they are buying products and services. He explained that it is important for procurement teams and stakeholders to have clarity as to why purchasing must be done a certain way. ExxonMobil has various types of businesses, including fuel and chemical businesses. End users from these businesses often collaborate with one another and procurement teams to determine whether they should buy a particular product, such as a

valve, on a micro-scale or at an enterprise-level. ExxonMobil's senior procurement leader told us the company's procurement team managers understand they are expected to meet the needs of the business units to help create value for shareholders.

- **AT&T** procurement leaders told us they collaborate with internal stakeholders to establish goals and metrics for their procurement teams that support the company's strategic goals. For instance, in pursuit of a strategic goal for revenue growth, the procurement leaders worked closely with the sales and logistics teams to establish metrics for measuring availability and turnover of inventory.

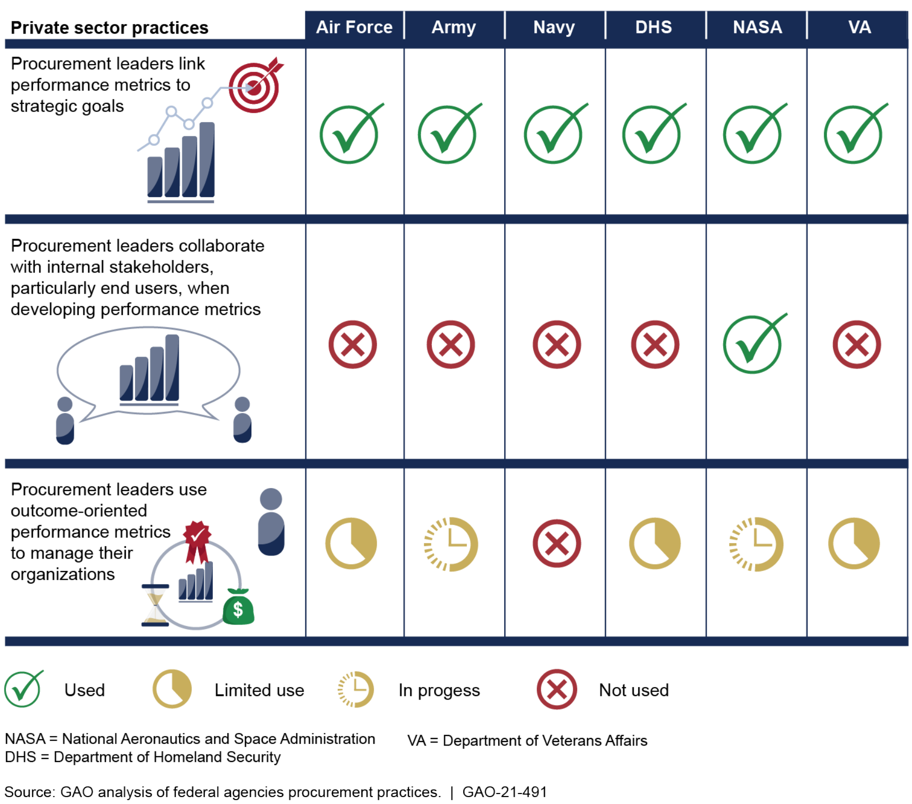
Procurement Leaders Use Outcome-Oriented Performance Metrics to Manage Their Organizations

Experts at academic institutions and professional associations told us companies use outcome-oriented performance metrics to enhance procurement operations. They said companies use these metrics to identify which of their procurement teams are achieving desired outcomes, such as reducing costs and improving performance. The corporate procurement leaders we interviewed emphasized the importance of using four types of outcome-oriented metrics: (1) cost savings/avoidance, (2) timeliness of deliveries, (3) quality of deliverables, and (4) end-user satisfaction. Corporate procurement leaders provided specific examples of how they have used outcome-oriented performance data to make management decisions, including the following.

- **Facebook** procurement leaders told us they use outcome-oriented performance metrics to identify operational deficiencies and make needed improvements, and that these metrics measure (1) whether products and services cost more or less than they should, (2) on-time deliveries, and (3) failure rates indicating the quality of deliverables. They emphasized that Facebook is a metric-heavy organization, and that it would be counter to their operating model to make decisions in the absence of outcome-oriented metrics. They also provided an example of how they used outcome-oriented procurement data. During a performance assessment, procurement leaders found some groups were missing performance targets for quality and timeliness of deliveries. This discovery drove additional analysis, and the procurement team determined that the lack of a dedicated contract execution team was contributing to these issues. After various discussions, company leadership created a contract execution team to help improve the quality and timeliness of deliveries.
- **Kroger's** senior procurement leader told us Kroger uses data to demonstrate how procurement teams are benefitting the company. As part of this effort, the leader told us he uses outcome-oriented metrics to measure cost savings, timeliness of deliveries, and quality of deliverables. He also said that prior to having a strong procurement organization with reliable outcome-oriented performance data, many management decisions were based on "I think, I feel, and I want," which led to poor decisions.
- **Raytheon Technologies** procurement leaders told us they have several outcome-oriented metrics, which measure cost savings, delivery times, and supplier quality. A senior procurement representative told us that these metrics are reviewed regularly to determine how the procurement organization is performing.
- **AT&T** procurement leaders told us they used end-user survey data to adjust their procurement team's buying behavior to better meet the company's needs. Specifically, they told us AT&T relies on real-time data and feedback from its frontline employees to adjust and optimize productivity while focusing on continuous process improvements.

Procurement Leaders at Selected Federal Agencies Did Not Consistently Use Key Practices to Improve Organizational Performance

Procurement leaders at all six of the federal agencies we reviewed linked the performance metrics for their procurement organizations to their agency’s strategic goals to help procurement personnel see connections between their daily activities and their agency’s mission. However, procurement leaders at five of the six agencies we reviewed told us that they did not collaborate with end users to develop performance metrics. Procurement leaders told us that they did not collaborate with end users for various reasons, including that end users were not particularly interested in the types of process-oriented metrics used to assess procurement organizations, and that they could define metrics appropriately without formal end-user input. As a result, the leaders missed opportunities to increase the usefulness and use of performance information in program management and policy. Additionally, the procurement leaders’ use of outcome-oriented metrics was limited, as they primarily relied instead on process-oriented metrics. The leaders cited various reasons for not using more outcome-oriented metrics. For example, two leaders told us that their current performance data for product and service quality are unreliable. As a result, the leaders lack balanced sets of performance measures that include both process- and outcome-oriented metrics, which we previously found help federal agencies identify improvement opportunities, set priorities, and allocate resources (GAO, 2015). Figure 4 shows the extent to which federal agencies used leading companies’ key practices when managing their procurement organizations.



Note: Our assessment of procurement leaders’ collaboration when developing performance metrics reflects the extent to which they collaborated with end users.

Figure 4: Federal Agencies Consistently Used One of the Three Key Practices Leading Companies Use When Managing Their Procurement Organizations

Procurement Leaders at Federal Agencies GAO Reviewed Linked Performance Metrics to Strategic Goals

Procurement leaders at all the agencies in our review linked their performance metrics to their agencies' strategic goals. These leaders stated that doing so helps ensure acquisition personnel are focused on the right things to support their agency's mission. These statements are consistent with statements we heard from procurement leaders at leading companies. Additionally, our previous work on creating a results-oriented culture found that explicit alignment between individuals' daily activities and organizational goals encourages individuals to focus on how they can help achieve those goals (GAO, 2003, 2017).

Procurement Leaders at Most of the Federal Agencies Reviewed Did Not Collaborate with End Users When Developing Performance Metrics

Procurement leaders at the six agencies in our review derived their performance metrics from statute, federal regulations, and OMB metrics, and collaborated with other members of the procurement community to develop performance metrics. However, only the procurement leaders from **NASA** collaborated with end users from the installation centers—such as technical experts—when developing performance metrics. In early 2021, NASA procurement leaders collaborated with these end users to develop a survey tool that collects quantitative and qualitative information that reflects end-user priorities. For example, the survey tool asks end users to rate the extent to which procurement officials met their needs, and gauges end-users' satisfaction with procurement officials' communication. NASA officials told us they initiated development of the survey tool in an effort to develop an end-user satisfaction performance metric after we provided them information about the key practices corporate leaders use in August 2020.

Procurement leaders at the other five agencies we reviewed did not collaborate with end users when they developed their performance metrics.

- The senior procurement leader at the **Air Force** told us he did not collaborate with end users, such as wing commanders, when developing performance metrics because he did not want end users to influence contracting operations excessively. He said too much end-user influence could discourage contracting officers from being business leaders, and lead to suboptimal results, such as narrowly pursuing specific, less-innovative solutions from industry. Instead, he collaborated with subordinate procurement leaders and members of the Air Force's financial management community to develop performance metrics for cost savings and cost avoidance, among other things.
- The senior procurement leader at the **Army** said she did not collaborate with end users, such as brigade commanders, and instead collaborated only within the procurement community when developing performance metrics.
- Procurement leaders at **Navy** headquarters, including the Navy's senior procurement leader, told us they did not collaborate with end users, including those representing the fleets, but that individual HCAs did collaborate with end users to develop performance metrics for their individual areas of responsibility. However, as a result, the performance metrics the Navy's senior procurement leader uses to assess activities across the entirety of the Navy are not informed by end-user input.
- **DHS's** senior procurement leader told us she did not collaborate with end users or their representatives when she developed performance metrics. She explained that end-user representatives, such as the leadership of the Border Patrol, are not particularly interested in the types of process-oriented metrics DHS uses to assess procurement

organizations, including small business utilization rates and workforce certifications. Instead, the DHS's senior procurement leader collaborated with other members of the procurement community, such as HCAs and small business proponents, to refine performance metrics and the associated targets. For example, for Fiscal Year 2020, she worked with the HCA for Immigration and Customs Enforcement to revise the component's "spend under management" target from $592 million to $779 million based on the prior year's performance.[40]

- In March 2021, **VA** procurement leaders told us their current performance metrics were not informed by end user input, but that they were in the process of testing a new survey tool they may use in the future, and that they had solicited end-user input through contracting officer representatives as part of that testing.[41]

Procurement leaders in the private sector told us that collaboration with end users during the development and implementation of performance metrics increases coordination and improves performance at the strategic level. While one senior procurement leader told us that too much end-user influence could lead to suboptimal results, leaders do not have to cede control when they collaborate with end users, and we have previously found that obtaining stakeholder input can increase the usefulness and use of performance information in program management and policy (GAO, 2005, 2013).

Procurement Leaders Primarily Rely on Process-Oriented Metrics, but Most Have Plans to Use at Least One Outcome-Oriented Metric In the Future

Procurement leaders at most of the agencies we reviewed have ongoing or planned efforts to use performance metrics to measure at least one procurement outcome, such as cost savings and end-user satisfaction. However, we found the leaders at all six of the agencies we reviewed rely primarily on metrics measuring processes, as they have in the past. These metrics are largely derived from OMB or statutorily required goals, and measure competition rates, small business utilization, and workforce certifications, among other things.

These leaders cited various reasons for not implementing metrics that are more outcome-oriented. For example, one leader said that too much focus on end-user satisfaction is a risk because some of the procurement community's innovations are achieved by focusing on mission rather than end-user satisfaction. While we recognize that too much end-user influence can introduce risk, corporate procurement leaders we interviewed emphasized the importance of measuring end-user satisfaction as part of their efforts to improve the performance of their procurement organizations. Additionally, two leaders told us current performance data for product and service quality are unreliable, although half of the leaders in our review are working to improve the quality of this data at their respective agencies, which can help facilitate the use of outcome-oriented metrics. Private sector procurement leaders we interviewed also noted they made concerted efforts to improve the quality of high-priority data.

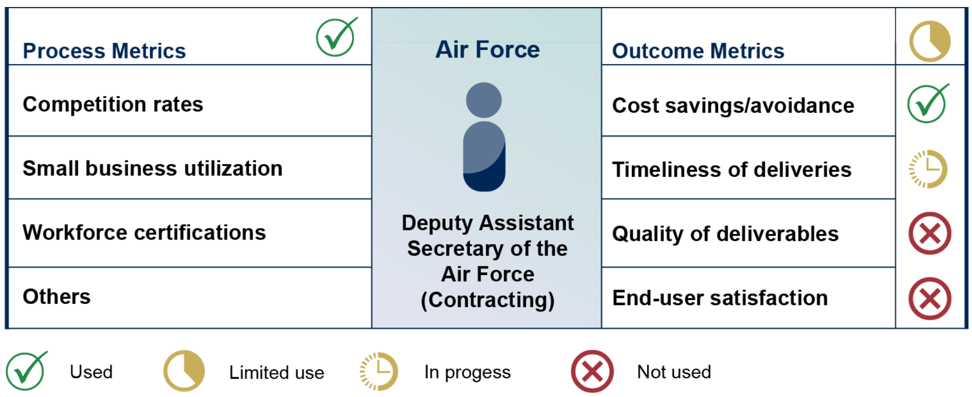
[40]The Office of Management and Budget uses a spend under management model to identify contracts that adhere to category management principles. When contracts adhere to those principles, that spending is considered "managed." According to the General Services Administration, increasing spend under management will decrease costs, contract duplication, and inefficiency; and lead to better buying outcomes. For additional information on spend under management, see GAO, *Federal buying power* (2020a).

[41]Contracting officers representatives are designated and authorized in writing by the contracting officer to collaborate with requiring activities and contractors to perform specific technical or administrative functions (FAR 2.101).

We have previously reported that establishing a balanced set of performance measures, including both process- and outcome-oriented measures, and obtaining complete and reliable performance information can help federal agencies identify improvement opportunities, set priorities, and allocate resources (GAO, 1996, 2005, 2015). Below, we assess the extent to which senior procurement leaders at the agencies we reviewed used the four types of outcome-oriented metrics used at leading companies: (1) cost savings/avoidance, (2) timeliness of deliveries, (3) quality of deliverables, and (4) end-user satisfaction.

Department of the Air Force

The Air Force's senior procurement leader has used a cost savings/cost avoidance metric to manage the Air Force's procurement organizations and is working to develop an outcome-oriented timeliness metric to supplement existing process-oriented metrics (see Figure 5).



Source: GAO analysis of Air Force documentation. | GAO-21-491

Figure 5: Air Force Performance Metrics for Procurement Organizations

The Air Force's senior procurement leader established a cost-savings tracker to identify, track, and validate **cost savings and avoidance** across the department. As of March 2021, the Cost Savings Tracker had identified $2.38 billion in cost savings and avoidance.[42] For example, the Cost Savings Tracker identified $158 million in cost savings for IT, which the Air Force achieved by adjusting IT refresh rates so they were driven by need rather than funding availability. Air Force procurement leaders told us they used the Cost Savings Tracker to identify additional opportunities to save or avoid costs, including by reassessing IT refresh rates. Additionally, the Air Force's senior procurement leader told us he was taking initial steps to obtain congressional approval for a pilot program to reinvest some of these savings back into the Air Force, which he said will incentivize decision-makers to reduce costs.

The Air Force's senior procurement leader is also working to develop an outcome-oriented metric for **timeliness of deliveries**, which he designated as Total Acquisition Lead Time (TALT). The Air Force defined *TALT* as the time from the identification of a requirement to

[42]As part of this process, the Air Force has standardized definitions for cost savings and avoidance. It has defined *cost savings* as reductions to budget lines or funded programs resulting from a new policy, process, or activity with no adverse impact on mission. It has defined *cost avoidance* as reductions in (1) the need for increased funding if present management practices continued; (2) unfunded requirements that were avoided; and (3) productivity gains, such as a reduction in required labor hours.

the delivery of a capability. The Air Force's senior procurement leader told us the Air Force does not currently have the technical capability necessary to measure TALT, but there are efforts underway to develop this capability.

In addition to these outcome-oriented metrics, the Air Force's senior procurement leader told us the Air Force currently uses **process-oriented** metrics to manage procurement organizations. For example, the Air Force assessed cycle-time data for sole source contract awards and identified factors contributing to longer cycle times. To address these factors, the Air Force's senior procurement leader deployed the DoD's sole source streamlining toolbox, which identifies actions procurement personnel can take to reduce cycle times and award these contracts faster.

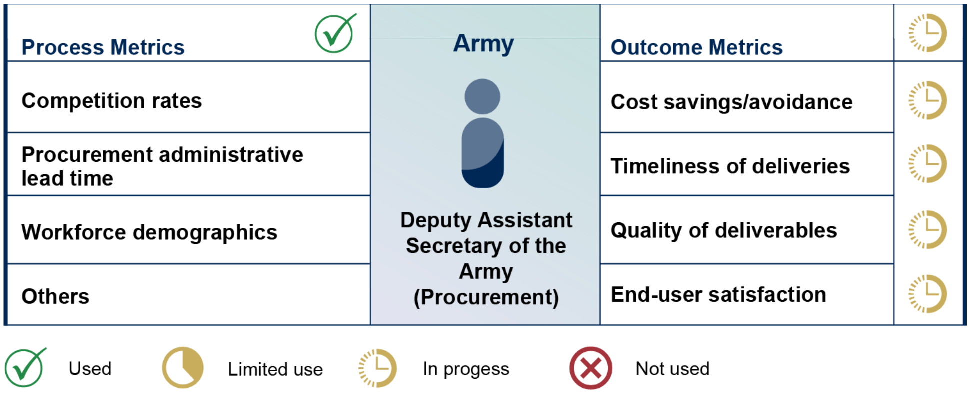
However, the Air Force's senior procurement leader has not pursued metrics to assess **end-user satisfaction** and the **quality of deliverables**. The official said this is because opinions vary about what end users should expect from procurement organizations, and what constitutes "quality" products and services. He also said that too much focus on end-user satisfaction is a risk because some of the procurement community's innovations are achieved by focusing on mission rather than end-user satisfaction. While we recognize that too much end-user influence can introduce risk, corporate procurement leaders we interviewed emphasized the importance of measuring end-user satisfaction as part of their efforts to improve the performance of their procurement organizations.

In addition, the Air Force's senior procurement leader told us he considers existing data on the quality of deliverables, specifically data in the Contractor Performance Assessment Reporting System (CPARS), to be generally unreliable.[43] This is a common challenge, and half of the leaders in our review are working to improve the quality of CPARS data at their respective agencies—which would facilitate more reliable outcome-oriented assessments. Private sector procurement leaders consistently told us it is important to measure the quality of deliverables, and that they make concerted efforts to improve the quality of important data. By using additional outcome-oriented metrics to assess the quality of deliverables and end-user satisfaction, the Air Force's senior procurement leader would have a more balanced set of performance measures to help identify improvement opportunities, set priorities, and allocate resources.

Department of the Army

The Army's senior procurement leader is developing several outcome-oriented metrics to supplement existing process-oriented metrics in an effort to field a balanced set of performance measures and better manage the Army's procurement organizations (see Figure 6).

[43]The contractor performance evaluation contained in CPARS is a method of recording contractor performance as required by FAR 42.15 and is used in source selection evaluations as required by FAR Part 15.



Source: GAO analysis of Army documentation. | GAO-21-491

Figure 6: Army Performance Metrics for Procurement Organizations

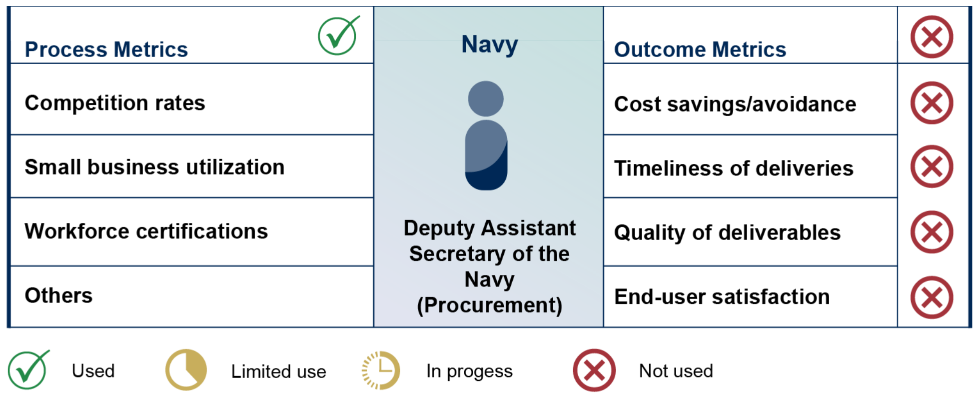
In the past, Army leadership reviewed **process-oriented metrics** at quarterly meetings to identify challenges and opportunities for improvement. For example, the Army's quarterly assessments of procurement administrative lead time helped the Army update workforce certification programs to provide acquisition personnel the knowledge needed to shorten lead times. The official told us the Army will continue to use process-oriented data to inform management decisions in the future, but procurement officials told us she suspended the quarterly review in 2020 when she started modifying the reviews to incorporate outcome-oriented metrics, among other things. As of March 2021, the Army had not yet determined a date for resuming quarterly reviews.

The Army's senior procurement leader has proposed using **outcome-oriented metrics** that match the types of outcome-oriented metrics commonly used by procurement leaders in the private sector: (1) negotiated **cost savings**, (2) **timeliness of deliveries**, (3) **quality of deliverables**, and (4) **end-user satisfaction**. The Army's senior procurement leader told us she began to pursue these outcome-oriented metrics in late 2020, after we provided her our interim assessment of the Army practices and how they differed from private sector practices. The Army's senior procurement leader envisions a dashboard where procurement organizations' performance can be viewed at any time, and plans to conduct reviews of outcome-oriented data on a quarterly basis. If the Army is able to develop this type of dashboard and consistently conduct quarterly reviews, it may provide the Army's senior procurement leader a balanced set of performance measures to help identify improvement opportunities, set priorities, and allocate resources.

In addition to performance metrics, the Army's senior procurement leader also uses Procurement Management Reviews to assess the health of the Army's procurement organizations. These reviews focus, in part, on workforce management and compliance with statutes and regulations, and they culminate in risk ratings for the Army's procurement organizations. The reviews are also a source for best practices, such as delivering training on key topics related to quality assurance. The Army's senior procurement leader is working to update the Procurement Management Review program to better align to procurement and Army-level strategic goals. The revised Procurement Management Review is intended to improve visibility into the Army's procurement organization's cost, schedule, and performance and identify any compliance problems.

Department of the Navy

The Navy's senior procurement leader has used process-oriented metrics to manage the Navy's procurement organizations and deferred responsibility for outcome-oriented performance assessments to the Navy's other HCAs, in accordance with direction from the Assistant Secretary of the Navy for Research, Development and Acquisition (see Figure 7).



Source: GAO analysis of Navy documentation. | GAO-21-491

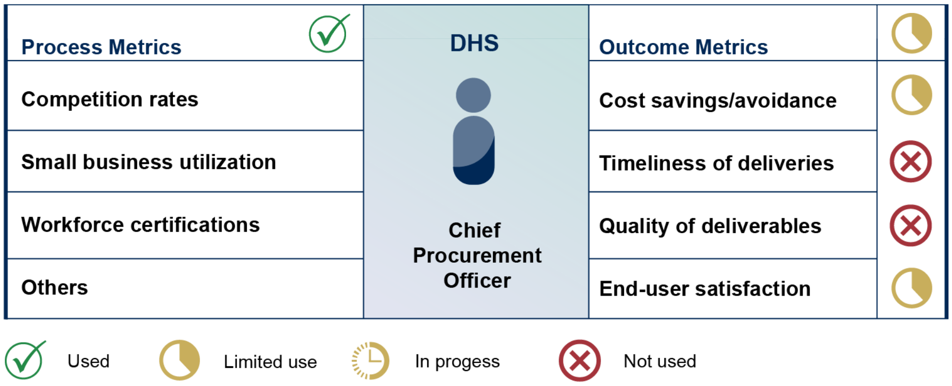
Figure 7: Navy Performance Metrics for Procurement Organizations

The Navy's senior procurement leader told us she has implemented a centralized approach for using **process-oriented metrics**—such as competition rates and small business utilization—across the department. The official told us that centralized process-oriented assessments provide broad visibility into HCAs' procurement processes and facilitate assessments of how well department-wide procurement processes are working. Additionally, the Navy's senior procurement leader uses the Navy's Procurement Performance Management Assessment Program, which primarily involves HCA self-assessments of procurement processes. These self-assessments are reviewed by senior Navy procurement personnel and subject matter experts to identify challenges, good practices, and lessons learned, which the Navy's senior procurement leader disseminates through a yearly newsletter. The Navy's senior procurement leader also participates in the Navy's "Two-Pass Seven-Gate" process, which the Chief of Naval Operations, Commandant of the Marine Corps, ASN (RDA), and other Navy leaders use to make investment decisions for large system acquisitions. The Navy's senior procurement leader told us this process provides her opportunities to influence procurement outcomes on a case-by-case basis.

The Navy's senior procurement leader told us she has not developed **outcome-oriented metrics**. The official defers outcome-oriented performance assessments to the Navy's 10 other HCAs, who have developed metrics unique to their organizations, because Navy leadership uses a decentralized approach to manage the department's various commands. However, a decentralized approach does not preclude the senior procurement leader from using outcome-oriented performance metrics in the same manner she uses process-oriented performance metrics. By using outcome-oriented performance metrics, the Navy's senior procurement leader would have a balanced set of performance measures to help identify improvement opportunities, set priorities, and allocate resources.

Department of Homeland Security

DHS's senior procurement leader has routinely used process-oriented performance metrics to manage DHS's procurement organizations, and has used end-user satisfaction and cost savings metrics on a limited basis. However, DHS's senior procurement leader has not used other outcome-oriented performance metrics (see Figure 8).



Source: GAO analysis of Department of Homeland Security (DHS) documentation. | GAO-21-491

Figure 8: Department of Homeland Security Performance Metrics for Procurement Organizations

DHS's senior procurement leader routinely uses **process-oriented metrics** to manage HCAs' performance in terms of competition rates, small business utilization, and other process-oriented activities. DHS's senior procurement leader regularly meets with the department's HCAs and reviews plans detailing the actions the HCAs intend to take to meet their targets for these process-oriented activities. DHS's senior procurement leader and HCAs also review policies, procedures, and training courses to identify additional opportunities to improve procurement processes. In addition to the metrics targeting the HCAs' performance, DHS's senior procurement leader uses metrics to assess industry engagement, and the extent to which contracting organizations value innovation, human relations, and other organizational traits.

DHS also measures cost savings achieved through category management activities, which are intended to improve how agencies procure common products and services, such as office supplies and building maintenance support. DHS officials told us the department's spending on common products and services accounted for about half of DHS's total contract obligations in Fiscal Year 2019 ($8.9 billion of $17.6 billion), and that the department saved $601 million through category management activities that fiscal year.[44]

However, DHS's senior procurement leader told us it would be difficult to identify cost savings for the remainder of the department's contract obligations because of unreliable data. DHS's senior procurement leader explained the department could compare actual contract costs to independent government cost estimates, but the quality of independent government cost estimates is inconsistent. DHS's senior procurement leader told us the estimates are often set to match the funding level the department has set aside for the contract in the budget, for example—a point that is consistent with our prior findings, where agency officials told us some independent government cost estimates were dictated by budget. As a result, DHS does not currently account for a large portion of its contract obligations when it calculates cost savings. Despite the challenge, corporate procurement leaders consistently told us it is important to measure cost savings/avoidance, and that they make concerted efforts to improve the quality of important data.

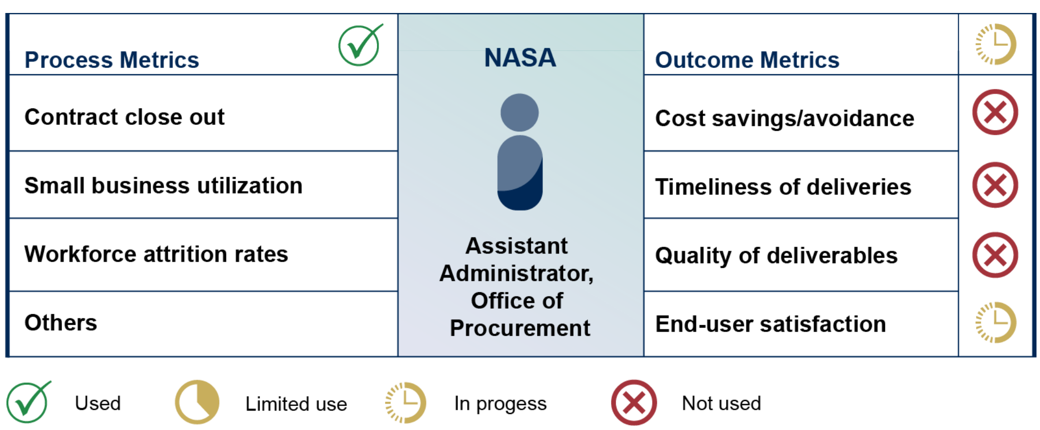
[44]In November 2020, we recommended OMB report cost savings from the category management initiative by agency, and OMB concurred with the substance of our recommendation. See GAO, *Federal buying power* (2020a). We are continuing to track OMB's actions in response to this recommendation.

Similarly, DHS's senior procurement leader has used an end-user satisfaction metric, but on a limited basis. DHS uses Acquisition 360 reviews to obtain feedback from stakeholders involved in a procurement, including contracting officer representatives. The department uses this process to review 100 contracts per year, assessing feedback on all aspects of the contracting process and identifying opportunities to improve operations. Furthermore, DHS established the Procurement Innovation Lab in 2015 to explore innovative procurement techniques, such as streamlined contracting approaches, and to share lessons learned—based in part on end-user feedback—among DHS's procurement community. In turn, DHS has disseminated the lessons learned to at least 1,750 personnel through training courses and other means. However, lessons learned through the Procurement Innovation Lab are based on a relatively small number of contracts. At the end of Fiscal Year 2019, DHS had awarded a total of 50 contracts through the Lab, but DHS awarded more than 23,800 contracts in Fiscal Year 2020 alone.

DHS's senior procurement leader told us the department does not have outcome-oriented metrics for the **timeliness of deliveries** and **the quality of deliverables** in large part because DHS lacks reliable data for these types of performance metrics, and applying the metrics to unreliable data would produce misleading results. For example, to measure timeliness and quality, DHS's senior procurement leader told us the department could attempt to use CPARS data, but it is challenging to ensure the quality of these data. This statement is consistent with the Air Force's senior procurement leader's position about CPARS data for Air Force organizations. Nonetheless, DHS's senior procurement leader, and leaders at the Army and VA, are working to improve the quality of CPARS data at their respective agencies, which would facilitate more reliable outcome-oriented assessments. Private sector procurement leaders consistently told us it is important to measure the timeliness and quality of deliverables, and that they make concerted efforts to improve the quality of important data. By using outcome-oriented metrics, DHS's senior procurement leader would have a balanced set of performance measures to help identify improvement opportunities, set priorities, and allocate resources.

National Aeronautics and Space Administration

NASA implemented process-oriented performance metrics across its procurement organizations, but NASA's procurement leaders have not yet used outcome-oriented performance metrics to manage NASA's procurement organizations (see Figure 9).



Source: GAO analysis of National Aeronautics and Space Administration (NASA) documentation. | GAO-21-491

Figure 9: National Aeronautics and Space Administration Performance Metrics for Procurement Organizations

NASA's procurement leaders told us they have quarterly meetings with the procurement leaders at NASA's 11 installation centers to discuss their organizations' performance against NASA's process-oriented procurement metrics, which measure contract closeout rates, small business utilization, and other aspects of the procurement process. Through these meetings, NASA's procurement leaders determine what processes are working well and what processes they should revise. For example, they observed an increased use of contracts with undefinitized terms and determined that it was due to lengthy proposal evaluations. They took steps to make evaluations timelier, which reduced the use of such contracts. However, NASA's procurement leaders have not set annual goals for the HCAs at the installation centers. Instead, NASA's procurement leaders have focused on whether NASA's installation centers are collectively achieving agency-wide goals. NASA's procurement leaders told us this approach is consistent with NASA's recent efforts to increasingly manage procurement across the installation centers as a single enterprise. They added that NASA's procurement leaders consistently work to ensure procurement organizations are adhering to the FAR. For example, the FAR states that firm-fixed-price contracts should be closed within 6 months, and NASA's procurement leaders monitor how long it is taking installation centers to close firm-fixed-price contracts.[45]

In addition to performance metrics, NASA procurement leaders also use information collected and analyzed through procurement initiatives aimed at improving their procurement practices. For example, NASA established a Source Selection Capability Group—comprised of subject matter experts from different installation centers—that assessed delays contributing to longer procurement lead times, and developed standardized document templates to increase efficiencies.

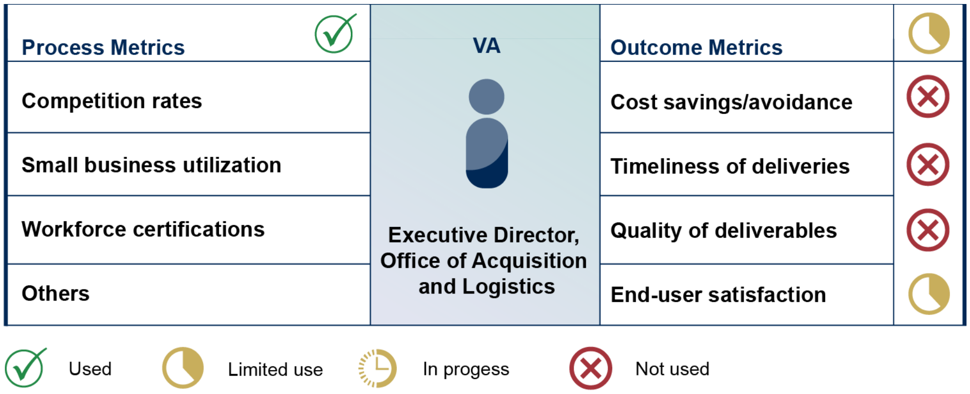
In August 2020, NASA's procurement leaders told us NASA was exclusively focused on implementing process-oriented performance metrics, rather than outcome-oriented metrics, because NASA's procurement leaders hoped process assessments and improvements would lead to better outcomes over time. As a result, NASA's procurement leaders do not have specific plans to use metrics measuring cost savings/avoidance, the timeliness of deliveries, or the quality of deliverables. However, in March 2021, NASA officials told us they had developed an end-user satisfaction survey in response to our interim assessment of their practices, that this survey is intended to help them develop end-user satisfaction metrics, and that they plan to start presenting the survey results during quarterly performance reviews by the end of Fiscal Year 2021. Using end-user satisfaction survey results in this way has the potential to facilitate more robust performance reviews, but NASA procurement leaders could better identify improvement opportunities, set priorities, and allocate resources by using a more balanced set of performance measures that include outcome-oriented metrics for cost savings and avoidance, timeliness of deliveries, and quality of deliverables.

Department of Veterans Affairs

VA's senior procurement leader has used process-oriented performance metrics to manage procurement organizations and implemented an end-user satisfaction metric, but lacks other outcome-oriented metrics (see Figure 10).

---

[45]FAR 4.804-1 stated that firm-fixed-price contracts, other than those using simplified acquisition procedures, should be closed within 6 months after the date on which the contracting officer receives evidence of physical completion. Contracts that require settlement of indirect cost rates should be closed within 36 months and other contracts should be closed within 20 months.



Source: GAO analysis of Department of Veterans Affairs (VA) documentation. | GAO-21-491

Figure 10: Department of Veterans Affairs Performance Metrics for Procurement Organizations

VA's senior procurement leader told us she primarily relies on process-oriented metrics—such as competition rates and small business utilization—to manage the department's procurement organizations. The official said these process-oriented data can help identify performance weaknesses and enable her to take corrective actions. For example, VA's percentage of competitive acquisitions receiving one bid ranked 21st out of 24 federal agencies in Fiscal Year 2018, and the VA's senior procurement leader told us she is currently working on collecting data to identify what factors have contributed to the department's low standing among other federal agencies with regard to this metric.[46] Further, the official has implemented an online knowledge portal and hosted acquisition innovation symposiums to help develop the department's procurement workforce. The VA's senior procurement leader is also co-chair of the Senior Procurement Council, which is composed of the department's HCAs and other internal stakeholders, such as small business proponents and attorneys. The VA's senior procurement leader told us the council meets at least quarterly to identify and address issues affecting the department's procurement organizations, such as the COVID-19 pandemic and the passage of the CARES Act.

The VA's senior procurement leader also piloted an Acquisition Management Review Program in Fiscal Year 2020 aimed at improving operations at VA's acquisition centers. Among other things, these management reviews included interviews with end-users such as clinical subject matter experts. For example, during the VA's National Acquisition Center management review, end users expressed concerns about poor communication with contracting staff. In response, the review team recommended that contracting officers hold regular meetings with end users.

Additionally, VA's senior procurement leader has undertaken efforts to improve the end-user satisfaction survey. For example, VA's senior procurement leader told us she is currently vetting potential updates to make the survey more useful for management decisions. In particular, VA's senior procurement leader told us she is expanding the scope of the survey to cover the entire acquisition life cycle, including requirements development and contract execution, since prior surveys focused solely on the contract award process. Further, VA's

[46]Competitive acquisitions receiving one bid refers to contracts awarded using competitive procedures for which only one offer is received (GAO, 2010).

senior procurement leader said she is broadening the survey's target audience in an effort to improve the survey's response rate of 14% in Fiscal Year 2019. The leader's efforts to improve the end-user satisfaction survey could help VA develop and use an end-user satisfaction metric.

VA's senior procurement leader, however, does not have similarly specific plans to improve visibility into cost savings/avoidance, the timeliness of deliveries, or the quality of deliverables, but did express a desire to eventually establish these types of metrics. Until these metrics are in place, VA's senior procurement leader will lack a balanced set of performance measures to help identify improvement opportunities, set priorities, and allocate resources.

Conclusions

There are inherent differences between the procurement organizations at federal agencies and leading companies. For example, procurement leaders at leading companies often focus on profit margins and return on investment, while procurement leaders at federal agencies do not. Additionally, procurement personnel at federal agencies are subject to laws and regulations intended to promote transparency and fairness, and to support socioeconomic goals. That said, there are also significant similarities between the procurement organizations at federal agencies and leading companies. Both buy a wide variety of critical products and services, and company leaders are expected to be good custodians of shareholder funds in the same way agency leaders are expected to be good custodians of federal funds. As such, there are opportunities for agency leaders to improve their organizations' performance by using some practices commonly employed by company leaders.

Unlike senior procurement leaders at leading companies, the senior procurement leaders at most of the federal agencies we reviewed did not collaborate with end users when they developed their performance metrics. While one procurement leader told us that too much end-user influence could lead to poor results, the leaders can collaborate with end users without ceding control to them. This type of collaboration increases buy-in from key stakeholders and the usefulness of the resulting performance information in management decision-making. Additionally, the leaders at the federal agencies did not routinely use performance metrics to measure key procurement outcomes, including (a) cost savings/avoidance, (b) timeliness of deliveries, (c) quality of deliverables, and (d) end-user satisfaction. Most of the leaders have plans to use some outcome-oriented measures in the future, and in certain instances they have taken the initial step of developing the metrics. However, they generally have not yet implemented the metrics in a routine or comprehensive manner, and two leaders said they had not done so because performance data for product and service quality were unreliable. Half of the leaders in our review were working to improve these data at their respective agencies, but currently the leaders' performance assessments face common limitations. They focus mainly on opportunities to improve procurement processes, while procurement outcomes receive less consideration. This imbalance is significant because the agencies' senior leaders use the assessments to set priorities and allocate resources intended to improve their organizations' performance.

Recommendations for Executive Action

Air Force

The Secretary of the Air Force should ensure the Deputy Assistant Secretary of the Air Force (Contracting): (1) collaborates with end users to develop performance metrics for procurement organizations; (2) uses a balanced set of performance metrics to manage the department's procurement organizations, including outcome-oriented metrics to measure (a) timeliness of deliveries, (b) quality of deliverables, and (c) end-user satisfaction.

Army

The Secretary of the Army should ensure the Deputy Assistant Secretary of the Army (Procurement): (1) collaborates with end users to develop performance metrics for procurement organizations; (2) uses a balanced set of performance metrics to manage the department's procurement organizations, including outcome-oriented metrics to measure (a) cost savings/avoidance, (b) timeliness of deliveries, (c) quality of deliverables, and (d) end-user satisfaction.

Navy

The Secretary of the Navy should ensure the Deputy Assistant Secretary of the Navy (Procurement): (1) collaborates with end users to develop performance metrics for procurement organizations; (2) uses a balanced set of performance metrics to manage the department's procurement organizations, including outcome-oriented metrics to measure (a) cost savings/avoidance, (b) timeliness of deliveries, (c) quality of deliverables, and (d) end-user satisfaction.

Department of Homeland Security

The Secretary of Homeland Security should ensure the DHS Chief Procurement Officer: (1) collaborates with end users to develop performance metrics for procurement organizations; (2) uses a balanced set of performance metrics to manage the department's procurement organizations, including outcome-oriented metrics to measure (a) cost savings/avoidance, (b) timeliness of deliveries, (c) quality of deliverables, and (d) end-user satisfaction.

NASA

The Administrator of NASA should ensure the NASA SPE uses a balanced set of performance metrics to manage the agency's procurement organizations, including outcome-oriented metrics to measure (a) cost savings/avoidance, (b) timeliness of deliveries, (c) quality of deliverables, and (d) end-user satisfaction.

Department of Veterans Affairs

The Secretary of Veterans Affairs should ensure the VA SPE: (1) collaborates with end users to develop performance metrics for procurement organizations; (2) uses a balanced set of performance metrics to manage the department's procurement organizations, including outcome-oriented metrics to measure (a) cost savings/avoidance, (b) timeliness of deliveries, (c) quality of deliverables, and (d) end-user satisfaction.

References

FAR 1.102-2, Performance Standards. https://www.acquisition.gov/far/1.102-2

FAR 1.601, General. https://www.acquisition.gov/far/1.601

FAR 2.101, Definitions. https://www.acquisition.gov/far/2.101

FAR 4.804-1, Closeout by the Office Administering the Contract. https://www.acquisition.gov/far/4.804-1

FAR 7.105, Contents of Written Acquisition Plans. https://www.acquisition.gov/far/7.105

FAR 15, Contracting by Negotiation. https://www.acquisition.gov/far/part-15

FAR 42.15, Contractor Performance Information. https://www.acquisition.gov/far/subpart-42.15

GAO. (1996). *Executive guide: Effectively implementing the Government Performance and Results Act* (GAO/GGD-96-118).

GAO. (2003). *Results-oriented cultures: Creating a clear linkage between individual performance and organizational success* (GAO-03-488).

GAO. (2004). *Best practices: Using spend analysis to help agencies take a more strategic approach to procurement* (GAO-04-870).

GAO. (2005). *Managing for results: Enhancing agency use of performance information for management decision making* (GAO-05-927).

GAO. (2010). *Federal contracting: Opportunities exist to increase competition and assess reasons when only one offer is received* (GAO-10-833).
GAO. (2013). *Program evaluation: Strategies to facilitate agencies' use of evaluation in program management and policy making* (GAO-13-570).
GAO. (2015). *Managing for results: Implementation of GPRA Modernization Act has yielded mixed progress in addressing pressing governance challenges* (GAO-15-819).
GAO. (2016). *Federal procurement: Smarter buying initiatives can achieve additional savings, but improved oversight and accountability needed* (GAO-17-164).
GAO. (2017). *Managing for results: Further progress made in implementing the GPRA Modernization Act, but additional actions needed to address pressing governance challenges* (GAO-17-775).
GAO. (2019). *Women-owned small business program: Actions needed to address ongoing oversight issues* (GAO-19-168).
GAO. (2020a). *Federal buying power: OMB can further advance category management initiative by focusing on requirements, data, and training* (GAO-21-40).
GAO. (2020b). *Small business contracting: Better documentation and reporting needed on procurement center representatives* (GAO-20-462).
Government Performance and Results Act (GPRA) of 1993, Pub. L. No. 103-62 (1993). https://www.govinfo.gov/link/statute/107/285
GPRA Modernization Act of 2010, Pub. L. No. 111-352, § 2(b). https://www.govinfo.gov/app/details/PLAW-111publ352
Services Acquisition Reform Act of 2003, Pub. L. No. 108-136, § 1421, 117 Stat. 1663 (codified as amended at 41 U.S.C. § 1702). https://www.govinfo.gov/content/pkg/PLAW-108publ136/pdf/PLAW-108publ136.pdf

PANEL 19. AGILE PRACTICES IN SOFTWARE-INTENSIVE PROGRAMS

Thursday, May 12, 2022	
11:05 a.m. – 12:20 p.m	**Chair: James L. Day Jr.,** Program Executive Office Integrated Warfare Systems, Director for Production, Deployment and Fleet Readiness ***Platform Infrastructure for Agile Software Estimation*** Nicholas Wagner, Institute for Defense Analyses ***Automated Data for DevSecOps Programs*** William Nichols, Carnegie Mellon University Hasan Yasar, Carnegie Mellon University Luiz Antunes, Carnegie Mellon University Christopher L. Miller, Carnegie Mellon University Robert McCarthy, Carnegie Mellon University ***New Effort and Schedule Estimation Models for Agile Processes in the US DoD*** Raymond Madachy, Naval Postgraduate School Wilson Rosa, Department of Homeland Security Cost Analysis Division Bradford K. Clark, Software Metrics Inc.

James L. Day Jr.— served the United States Navy in multiple capacities for over 27 years. He was selected to his present position and to the Senior Executive Service in August, 2020. As the Director, Production, Deployment and Fleet Readiness within the Program Executive Office Integrated Warfare Systems (PEO IWS), Mr. Day is responsible for the integrated production and fielding requirements for weapons, sensors, combat systems and associated system elements. Mr. Day leads the overall combat system, system engineering efforts that result in continuous improvement in production capability, capacity and maintenance of fleet readiness.

Prior to his present position, Mr. Day served as the Deputy Program Manager for the Guided Missile Frigate Construction Program (PMS 515). While serving as DPM, Mr. Day led the accomplishment of several major program milestones including the award of one of the largest full and openly competed ship construction contracts in Navy history. Previously, Mr. Day served on the staff of the Assistant Secretary of the Navy, Research, Development and Acquisition (ASN RD&A) providing oversight and guidance for all Navy, Above Water Sensors and Laser Weapon Systems. Mr. Day also served as the Principal Acquisition Program Manager for LCS Combat Systems, the lead for DDG 51 Combat System, Ship Integration and the lead for AN/SQQ-89(V) Advanced Development and Production in PEO IWS. He has also held multiple leadership positions at the Naval Undersea Warfare Center, Newport, RI. Mr. Day is a veteran of the United States Navy having served from 1992-1998 and achieving the rank of Sonar Technician, Second Class.

Mr. Day received his BS in Computer Information Systems from Roger Williams University and his Master of Business Administration from the University of Maryland Global Campus. Mr. Day is also a graduate of the Executive Program Managers Course at Defense Acquisition University and the Executive Leadership Certification Program from Cornell University.

Platform Infrastructure for Agile Software Estimation

Nicholas A. Wagner—has been a Research Staff Member at the Institute for Defense Analyses (IDA) since 2019. Before joining IDA, Nicholas was a materials science and engineering PhD student at Northwestern University, where he applied machine learning to the discovery of new electronic materials. While at Northwestern, he was a visiting student at Los Alamos National Lab as well as a research software engineering intern at Google Accelerated Science. His research interests at IDA include the application of data science techniques to cost estimation and program evaluation, crowd forecasting, and the governance of AI. Nicholas is a lifelong nerd and Phoenix native. [nwagner@ida.org]

Abstract

Software acquisition reform is a hot topic in the DoD, but the oversight community is struggling to adapt to changes. I diagnose some issues with the current state of business in the Software Acquisition Pathway and propose a system called Overlord to increase the level of automation in software program management and oversight. My goal is to make life easier for software developers, program managers, and members of the oversight community.

Introduction

Nearly 11 years after Marc Andreessen's claim that "software is eating the world" (Andreessen, 2011), the Department of Defense (DoD) is still struggling to appear appetizing. The department is well aware of its barriers to software acquisition from high-profile reports such as the National Security Commission on AI report in 2021 (Schmidt et al., 2021) and Defense Innovation Board (DIB) Software Acquisition and Practices (SWAP) study in 2019 (Defense Innovation Board, 2019). Actions are being taken to lower the barrier to acquiring software, such as developing the Software Acquisition Pathway (SAP); building continuous authority to operate DevSecOps platforms like the Navy's Black Pearl, the Army's CReATE, and the Air Force's Platform One; and establishing "software factories" within the services. These innovations have returned results real enough that the model is being matured for scale across the DoD. Defense Deputy Secretary Kathleen Hicks has called for tighter integration of service software factories on a "reasonable" number of service providers and software repositories (Serbu, 2022). Doing so will help control cloud service costs as well as reduce technical barriers to code reuse, bringing DoD software development practices across the enterprise closer in line to commercial ones.

This integration of developer-friendly platform infrastructure is a welcome advance, but there is one key group of users who need more: the oversight community. As acquisition processes change to more closely follow the commercial technology sector's agile approach, the way programs are assessed for cost, schedule, and value must evolve as well. And just as enterprise development infrastructure has enabled a software development revolution for the DoD, an enterprise infrastructure for oversight could also unlock great value for the oversight community. In this paper, I lay out a proposal for a data system I call Overlord to do just that. I begin with what I see as the main job roles affected by oversight of the SAP and what their desires are. I then discuss how the current state of oversight leaves those desires unsatisfied. To inform my proposed solution, I describe two examples of currently existing infrastructure, Platform One and the Cost Analysis Data Environment (CADE), and how they meet or fail to meet the desires of intended users. After, I suggest how the DoD can remedy these problems with an enterprise data system called Overlord. Finally, I sketch out some ideas for implementation.

The Roles

There are three archetypal job roles in the software acquisition world: developers, overseers, and program managers. Developers are the most straightforward in their desires. They want to build interesting software and be troubled as little as possible with busywork that detracts from that. This is an understandable instinct, but it can lead to phenomena such as "nerd sniping"—when developers become fixated on interesting subproblems to the detriment of the larger goal (Munroe, 2007)—or other behaviors that delay delivery to the end user. "Overseers" is a term used here as a catchall for people who are interested in making sure resources dedicated to a project would not be better used elsewhere. This group includes staff overseeing portfolios of projects, service and OSD cost estimators, inspectors general, the GAO, and the like. Overseers are also interested in a project's success, but they are equally, if not more, concerned with the use of taxpayer funds in a manner that minimizes the risk of waste. Due to their concern with risk, overseers typically desire as much information as possible as a condition of granting resources so they might better forecast which projects will fail. This desire naturally conflicts with developers' desires to focus on their work. In this scenario, program managers are placed in the middle. They focus on acquiring and keeping the resources they need to deliver their project in a manner satisfactory to the customer. This approach usually entails tracking critical metrics of value, assigning work to team members, pushing for adherence to the expected road map, holding meetings to get status updates from developers and remove their obstacles, and providing updates to overseers in turn. Great program managers shoulder these necessary responsibilities while keeping the overhead burden on their developers to a minimum. I summarize these three roles and their relationships in Table 1.

Table 12. Motivations and Risks for Software Acquisition Archetypal Roles

	Developer	***Program Manager***	***Overseer***
Primary Selfish Objective	Wants to minimize time not developing software	Preserves program resources, keeps developers producing	Wants as much information as possible to predict the future
Primary Risk to Organizational Success	Not always focused on risk to program portfolios or taxpayer	Overloads developers with reporting or fails to communicate crucial management info	Not always considerate of developer time pressures

The balance of this tripartite ecosystem has important consequences for the industrial base. Overly expansive documentation requirements for oversight are both literally costly in the sense that record systems and staff to run them cost money, but also in the sense that developers have nonmonetary concerns about how bureaucratic their work is. These concerns can drive off nontraditional vendors and talent. On the other hand, insufficient transparency can make it impossible to make informed assessments of program performance or estimates of future software cost, schedule, and quality. The optimal trade-off is getting the information needed to effectively oversee a software program to the right people with a minimum amount of effort on the part of developers. I discuss the SAP's attempt to thread this needle next.

The Current System of Software Program Oversight

The 2019 DIB SWAP study documented several deficiencies with how software programs were typically managed. Summarized, the DoD treated software like hardware, trapping software programs in a 1970s paradigm of waterfall development where speed took a back seat to years-long planning. Unscheduled change was viewed as a planning failure rather than responsiveness to user needs. Software was taking too long to acquire, cost overruns were common, and failures to adapt to changing requirements resulted in public embarrassments. The SWAP study called for an alternative to the traditional combination of Joint Capabilities Integration and Development System (JCIDS); Planning, Programing, Budget, and Execution (PPB&E); and Defense Federal Acquisition Regulation Supplement (DFARS). This resulted in the Software Acquisition Pathway (SAP) a year later. SAP programs will be the focus of this paper as they represent the current state of the art when it comes to agile software development.

SAP program reporting is rooted in some of the SWAP study's other recommendations. As part of its work, the SWAP study developed a guide on "Metrics for Software Development" (DIB, 2019, S88-S90). It concluded that the use of source lines of code (SLOC) as a software complexity metric and SLOC/unit time as the corresponding productivity measure for developers was not a productive way to measure the status of programs. The guide dived deep into suggested program metrics covering deployment rates, response rates, code quality, functionality, and importantly to the overseers, software program management, assessment, and estimation. The full list is not pasted here as it is fairly lengthy, but I recommend reading either the source material or the condensed list provided on the Defense Acquisition University's (DAU) page on the SAP's Program Management Metrics and Reporting (DAU, 2022).

The exact reporting requirements for SAP programs are described next. Discussed first is the simplest case of those with less than $20 million of software expenditures and below the Acquisition Category (ACAT) 2 threshold. Currently, the primary recipients of metrics reporting for such programs are program managers, their immediate decision authority (DA), and the Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD(A&S)).

Small Program Oversight Data

In terms of key data collection, program managers or their end-user sponsors are required to collect four main sets of quantitative information: value assessments, cost estimates, program metrics, and road maps. Value assessments are conducted annually at a minimum and capture outcome-based measures of a program's ability to meet end-user mission goals. Cost estimates are a responsibility of program managers and must be prepared before program execution and updated at least annually. Additionally, program managers must develop a Metrics Plan. The Metrics Plan identifies metrics to be collected in order to manage the software program's day-to-day performance. This plan may overlap with some of the information collected for the value assessment, but it should also contain information on process efficiency, software quality, software development progress, cybersecurity, and cost. A minimal subset of these metrics is required to be provided quarterly to the relevant OUSD, such as OUSD(A&S) or OUSD(R&E). Road maps show planned goals and features of each software iteration over the next 18 months. The OUSD(A&S) provides guidance on all four of these items, including recommending that programs provide data access to approved stakeholders with automated read-only self-service portals. However, the exact metrics and manner of metric collection and sharing is ultimately left up to the programs.

In addition to the previously mentioned requirements, some data explicitly used for SAP review and not program oversight must be reported to OUSD(A&S) semiannually. As of March 5, 2022, the reporting guidance on the DAU's "Program Management Metrics and Reporting" page tells decision authorities of programs using the SAP to email a form to OUSD(A&S) with 11 metrics twice a year (DAU, 2022). The use of email is intended to be temporary while formal reporting systems are updated. Those metrics are:

1. Avg Lead Time for Authority to Operate (days)
2. Continuous Authority to Operate In-Place
3. Mean Time to Resolve Experienced Cyber Event

4. Mean Time to Experience Cyber Event
5. Avg Deployment Frequency
6. Avg Lead Time
7. Minimum/Maximum Lead Time
8. Avg Cycle Time
9. Change Fail Rate
10. Mean Time to Restore
11. Value Assessment Rating

Larger Program Oversight Data

For programs above $20 million in software expenditures, a Contractor Cost Data Report (CCDR) and Software Resources Data Report (SRDR) must also be filed. CCDRs and SRDRs are filed at program start and then regularly with each major release until program completion. They are Excel or JSON files containing information on expected and actual software size and complexity, cost breakdown, and development schedule. These data are fed into the Cost and Software Data Reporting (CSDR) system managed by the CADE under OSD's Cost Assessment and Program Evaluation (CAPE) office. For programs above the ACAT 2 threshold, an independent cost estimate (ICE) by CAPE is required. CAPE analysts will therefore directly receive metrics in some manner and conduct site visits. CAPE is required to be notified 210 days before the execution decision is made and must hold a kickoff meeting no later than 180 days before the execution decision.

Problems with the Current System

There are three main problems with the current system:

1. Data that would be useful across the enterprise is not available at the enterprise level.
2. Oversight activities take a long time, limiting a program's ability to rapidly execute.
3. Automation is an additional responsibility.

The first point is a consequence of the high level of delegation to program managers for most information collection. Aside from the 11 metrics required by OUSD(A&S) semiannually, the program's metrics are collected in whatever manner the program manager decides. OUSD(A&S)'s system for storing and organizing the information reported to them is unclear. For larger programs, some information is collected within CADE in the SRDRs and CCDRs, but these reports only partially cover information from the value assessments and program metrics. And when CAPE conducts a cost estimate, its analysts are free to collect additional data in yet another manner. As a result, the ability to access and analyze historical program data is compromised. These varied approaches make it difficult for other programs to improve the accuracy of their own estimates, more difficult to assess past performance, and more difficult to develop tools for process automation (Putnam-Majarian & Staiger, 2019). This lack of access to data is reflected in the rather sparse published studies looking at agile software estimation in the DoD and their small sample sizes (Goljan, 2021; Madachy, 2018).

The second point on long execution times for oversight is largely a result of the first, but lack of historical data is not the sole cause. There is also the issue of a lack of tooling supporting common activities like cost, schedule, and value estimation. Because of difficulty locating and obtaining relevant program data and reliance on manual analysis, cost estimators take a long time to prepare their analyses. Furthermore, the estimates themselves are preserved in a manner inconducive to automated updating or reuse (e.g., as Excel files with expert knowledge embedded in their construction). In a way, the situation is actually reminiscent of the modelling challenges facing the Joint Artificial Intelligence Center (JAIC), which also struggles with data access and model operationalization. I should also note that while the half-year minimum timeline for independent cost estimates is public, the timeliness of program cost estimates is unknown but presumably faster. In industry practice and under USD(A&S) guidance, cost/schedule/value estimates are iteratively refined as time goes on and more data is collected. In the commercial technology industry, estimates are updated in real time or, at the very least, at the cadence of a sprint cycle. If government estimates were compressed to this timescale, there would be much better visibility into problems before they occurred and less of a burden on programs when estimates need to be conducted.

The use of automation was recommended by the SWAP study authors as a solution to the timeliness issue and has since been repeated by other bodies such as the Government Accountability Office (GAO, 2021). This key role of automation was encoded in DoD Instruction 5000.87, Section 3.3.b.11, which demands that SAP programs' "metrics collection will leverage automated tools to the maximum extent practicable" (DoD, 2020, p. 18). But what makes automation practicable? Programs are already resource-constrained fulfilling the needs of their users. The addition of reporting automation development to their existing workload is unlikely to be prioritized.

These challenges are not being faced for the first time. Before proposing a solution, it is helpful to explore the examples of two systems already discussed in this paper, Platform One and CADE.

Infrastructure Examples: Platform One and CADE

Platform One

Platform One is an official DevSecOps Enterprise Services team for the DoD housed within the Air Force Office of the Chief Software Officer. Platform One provides several backend services for software developers in the DoD, such as prescreened software containers, cloud-native access points, identity management capabilities, automated testing pipelines, and coding collaboration tools. These services allow the operation of "software factories," which developers in program offices use to more easily produce operational software. Users are charged fees commensurate with their level of demand on the Platform One team. If no team labor is required, users frequently pay nothing. By turning industry-standard software development tooling, security, and operations into services, Platform One frees developers to stop repeating the 85% of work common to all DoD programs (Platform One, 2020), and lets them complete their program-unique work more quickly. A diagram (Figure 1) provided by Mr. Nicolas Chaillan, former Chief Software Officer of the Air Force, illustrates how Platform One serves as the connective layers between the cloud computing base infrastructure and the application layer being created by program software development teams (see "YOU" in Figure 1). Before the establishment of Platform One, developers with knowledge of industrial best practices knew these layers were critical, labor-saving capabilities in software engineering. However, each program was expected to pay for and stand up such capabilities on their own, much like how SAP programs today are expected to adopt automation in their oversight reporting. Platform One demonstrates the time and cost savings that can be realized by investing in enterprise-wide capabilities.

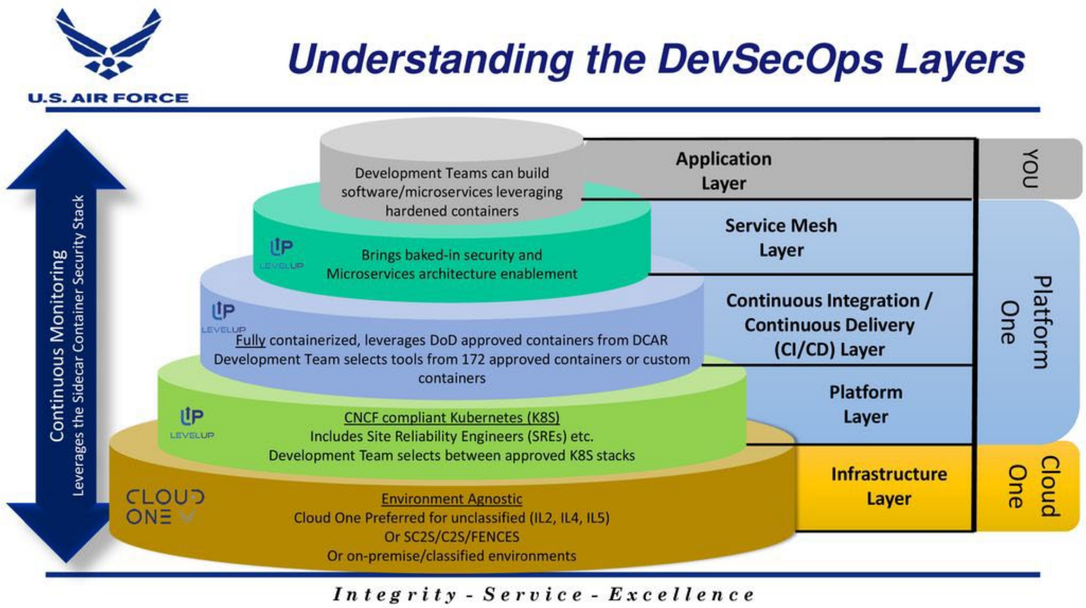


Figure 16. Services Provided by Platform One
(Chaillan, 2020)

CADE

The Cost Analysis Data Environment (CADE) is a data system run by OSD CAPE. It serves as the official repository for Contractor Cost Data Reports (CCDRs/1921s); Cost Analysis Requirements Descriptions (CARDs)/Technical Data (1921-Ts); Selected Acquisition Reports (SARs); and the Software Resource Data Reports (SRDRs) submitted by programs. CADE's vision is to "provide comprehensive data availability and automate common data visualization methods to help depict each program's unique story" (CADE, 2022c). Government overseers can search within their web browsers for data by several categories, including service, program, ACAT, contractor name, and a few others. Once data is located, CADE can provide visualizations and allow users to download the data, typically as an Excel file or PDF. CADE also provides a Data set, Tools, and Modelling Hub (DTMHub) where groups of users can share resources that they manage. These resources might take the form of an Excel spreadsheet that can be downloaded, a software application, or a link to another website. Note that there are some gaps in the data available in CADE on agile software projects. This data is generally more applicable to the waterfall method of software development historically prevalent in DoD software acquisition programs. Only in 2017 did the reporting of agile-focused alternatives to traditional software metrics, like lines of code, become possible on SRDRs, and their inclusion was not mandatory. As mentioned previously, programs with software efforts less than $20 million are not typically required to submit SRDRs, and therefore data on smaller software programs in the SAP is also sparse. Despite these limitations, CADE is better than the many stovepiped systems that it replaced. In recent years, CADE has added functionality for its users and updated documentation at a decent clip. Where CADE perhaps falls short, aside from its data availability issues, is its focus on manual workflows (Figure 2) via graphical user interface (GUI) rather than on automated ones.

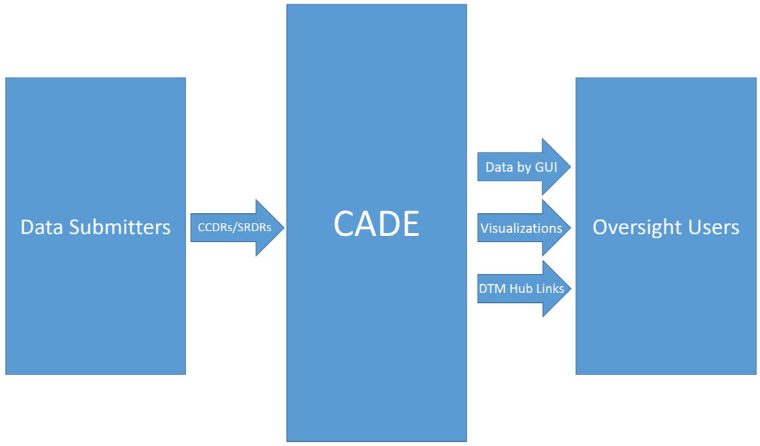


Figure 17. Sketch of CADE's Use for Software Programs

An Enterprise Infrastructure Proposal

I propose a potential enterprise infrastructure combining the commercial technology best practices of Platform One with the demonstrated oversight utility of CADE to address the challenges of data availability, speed, and automation raised previously. For convenience, I will call the hypothetical system "Overlord." A rough sketch for Overlord is presented in Figure 3, with its components discussed in the following sections.

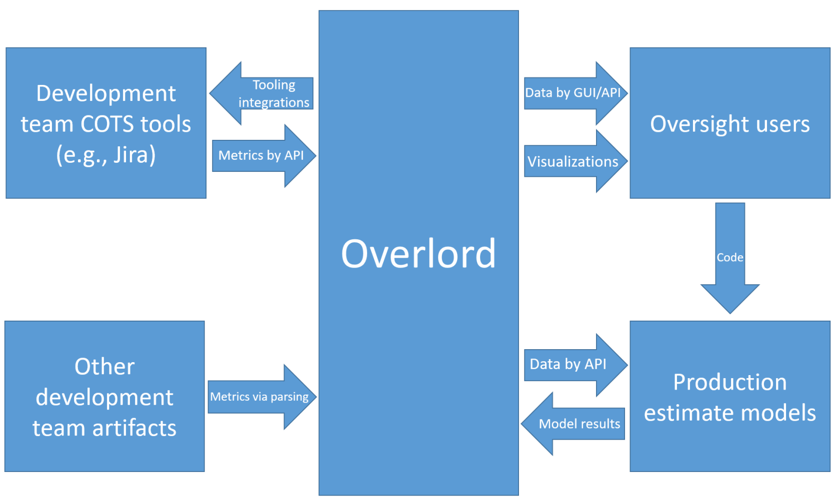


Figure 18. Overlord System Sketch

I begin with the challenge of automated data extraction. I suggest providing automation support for agile software program metrics in an iterative manner beginning with what are already known to be small lifts with high impact. As already known to readers who have utilized popular commercial off-the-shelf (COTS) software development tools such as GitLab, Bitbucket, or GitHub, these applications can instantly supply certain types of information such as commit activity data, team size, merge rates on pull requests, and other information with no manual effort through application programming interfaces (APIs). GitLab and other popular project management tools such as Jira and Asana can also export key agile metrics such as story point counts and backlog burndown rates. By leveraging existing automation and building out more capabilities to do so, effort required of developers can be minimized. Rather than have each program set up their own connections to these APIs to populate reports for emailing, Overlord would provide its own published API that handles receiving this information into a database. Publishing this API is important as it prevents the vendor lock associated with tying support to one particular tool. After an enterprise team writes the software to connect two APIs, developers on program teams would grant the specific access required to establish a link. The data could then be updated automatically moving forward with a simple script. Since many COTs tools are used by multiple teams, this setup can require development effort once and then be repeated easily, similar to how Platform One leverages commonalities among software development efforts. Of course, not every development team will use COTS solutions for some of these metrics, especially those contained in the value assessments, so Overlord would also support upload and parsing of workhorse formats such as Excel, CSV, PDF, Word, and JSON. This parsing will be much easier with provision of standard templates and detailed guidelines similar to those provided by CADE for SRDRs. Support for other tools will be added gradually over time in order of demand.

There is also some work to be done on common infrastructure regarding how data is pushed out, not just pulled in. CADE, for example, has a web browser GUI. Only within the last year was support added to download SRDR data in bulk (CADE, 2022b). APIs can help here as well, allowing various types of users to pull the exact data they need automatically without necessarily needing to understand the inner workings of Overlord's database. This would enable, among other things, easier integration with enterprise data platforms like Advana and live feeds to cost, schedule, and value models. Another big advantage of an API is low overhead for access role creation. For example, a contractor could be granted access to pull data only from their own past projects, or an academic researcher could be granted access only to non-proprietary fields or aggregated quantities. These types of access are currently forbidden in CADE, out of both a fear that proprietary data will be inappropriately accessed and the infeasibility of someone manually curating data for sharing.

Not that web browsers should be entirely neglected in favor of access by API. Once again, CADE has some features to highlight here. One is that users can search by certain structured fields like program, WBS, service, and so on. These ideas could be extended further by allowing for searches of unstructured full text using something like Solr or Elasticsearch. Such search functionality would be especially vital when many program documents contain vital information in narrative form. Another feature of CADE worth copying is the automatic generation and display of commonly used plots for traditional acquisition programs like average procurement unit costs over time. Overlord would supply similar functionality but with common software development-related plots like product road map, sprint burndown, epic burndown, velocity, and control charts. Users would be free to export the plot data for re-creation in their own preferred plotting tools, if desired.

But the collection, sharing, and display of program data is not the only purpose of this system. The data it collects would also be a key driver of enhancements in development productivity. For instance, one of the common complaints about agile software estimation is the inconsistency of software size measures, such as story points between teams or even the same teams over time. Story point planning is a manual process that is subject to biases and uses lots of personal experience and gut intuition. This is especially hard on newer teams and makes analogizing coding speed estimates to later projects difficult. By capturing historical schedule metadata, extensions for development planning tools like Jira can be built that show the historical point totals and schedule actuals for similar stories, reducing mental effort on the developer's part and increasing consistency. Similarly, teams often struggle with decomposing high-level features into a set of digestible, specific tasks. Overlord can provide suggestions for developers from close past examples, increasing the thoroughness of planning. As time goes by, the amount of data collected and feedback on suggestions would increase, improving the quality of the suggestions.

There has been plenty of discussion of how machine learning and AI will change the nature of estimating itself in the DoD. Perhaps bucking the trend, I do not suggest building another platform for hosting production machine learning models. I think it would actually be a better use of the oversight community's resources to piggyback on the JAIC's work to put models into operation and host estimation models on their infrastructure. Instead, a hub would be provided on Overlord where users could view model outputs with dashboards and obtain links to the data supplied from Overlord for the model, the model documentation, and code used. This would essentially be an expanded version of CADE's DTMHub.

To illustrate Overlord's potential, imagine the following demonstration:

Alice is a new cost analyst in CAPE. She has been instructed by an SES, Bob, to look at a months-old program to see whether its planned road map is realistic. She turns to a more experienced colleague, Candice, to ask her advice. Candice asks for the program's name and quickly looks it up in Overlord's web application. Seeing that the road map in Overlord looks outdated, she asks Alice whether she can get the latest copy. Alice copies a code example from Overlord's code hub, changes the "INSERT_API_KEY" value to her own, and modifies the project ID number. A few seconds after pressing run, the program's Jira instance returns a list of major stories and their expected dates formatted as a blob of JSON. A later section of Alice's code ingests the JSON file into Overlord, checks its validity, and puts it in Overlord's database. Candice refreshes the page, and a newly updated road map chart appears. She asks whether Alice has examined any prior programs run by the program's manager. Hearing no, Candice queries Overlord for past projects involving the program manager and filters for story point data. Downloading the data as an Excel file, Candice emails the data set to Alice and says she should take a look at what the projected and actual story point velocities were. Before Alice heads back to her desk, Candice also suggests she look at the schedule estimate from CAPE's baseline model for new programs. Candice shares a link on Microsoft Teams to the live model page, which provides a dashboard where users can input a program's parameters and see the predicted likelihood of meeting the schedule from a model retrained every day on fresh data. She notes to Alice that the program manager can see the same model on his program's Jira project page, so she shouldn't expect a lot of disagreement. Alice thanks Candice and walks away.

This vignette illustrates the potential of Overlord to turn time-consuming tasks like updating data, making plots, and sharing estimate models into quick, nearly effortless activities. Following these suggestions, costs to collect and share all kinds of information can be driven down to nearly nothing. This enables a win-win mindset where developers save time while receiving helpful guidance, and overseers get constant access to troves of useful data.

What It Will Take

Overlord can build upon past investments in oversight data systems. As mentioned before, CADE has established a data warehousing functionality for traditional CSDR and other non-software-related data. Some efforts to automate data collection are also underway with the CSDR Planning and Execution Tool (cPet) that creates report templates in Excel or JSON format and validates them when completed (CADE, 2022a). Extending the capabilities of CADE to handle what I envisioned previously is not a far stretch of the imagination. A question, then, is what role other organizations play. OUSD(A&S) owns the SAP and therefore has both a major stake in the outcome and a head start on managing SAP program data from their collection of semiannual reports. From an outsider's perspective, it seems intuitive that both organizations share a common interest in collaborating on a system like the one I propose. There is also the question of funding. One suggestion is to charge programs a small fee for using the system the way Platform One charges for managed services. The current alternative of emailing forms and waiting 210 days for an ICE would still be a free option. This approach would align the incentives of the Overlord team with the program team's desire for speed and ease of use compared to the alternative. The development of Overlord would require software engineering and data engineering talent like every other modern data initiative. Likely staffing partners include the Cost Estimating and Analysis Data Tools Tiger Team led by OSD CAPE for their data science and cost estimating expertise, as well as members of the service software factories who are well-acquainted with the state of the art in development software tooling. Like the SAP programs themselves, it is best to begin with a small team and follow the same general iterative steps (certified need statement, value assessments, road map, etc.) to build Overlord. With the right talent and development process, Overlord can quickly begin serving user needs in the agile oversight community.

REFERENCES

Andreessen, M. (2011, August 20). Why software is eating the world. *Andreessen Horowitz*. https://a16z.com/2011/08/20/why-software-is-eating-the-world/

Chaillan, N. (2020, September 10). *DoD Enterprise DevSecOps Initiative and Platform One*. https://www.youtube.com/watch?v=gR0Mzo7WDss&t=2s

Cost Assessment Data Enterprise (CADE). (2022a, March 6). *CSDR Tools*. https://cade.osd.mil/tools/csdr-tools

Cost Assessment Data Enterprise (CADE). (2022b, March 6). *News*. https://cade.osd.mil/News/blog#news_20211012

Cost Assessment Data Enterprise (CADE). (2022c, March 13). *About Us*. https://cade.osd.mil/about/cade

Defense Acquisition University. (2022, March 4). *Program management metrics and reporting*. https://aaf.dau.edu/aaf/software/metrics-and-reporting/

Defense Innovation Board. (2019, May 3). Software is never done: Refactoring the acquisition code for competitive advantage.

DoD. (2020, October 2). *Operation of the software acquisition pathway* (DoDI 5000.87). Department of Defense. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500087p.PDF

Goljan, J. (2021). *Developing a Cost of Delay (CoD) framework for the DoD & analyzing the current state of Air Force agile cost estimation* (AFIT-ENV-MS-21-M-231) [Master of Science in Cost Analysis]. Air Force Institute of Technology. https://scholar.afit.edu/cgi/viewcontent.cgi?article=5949&context=etd

GAO. (2021, September 30). *DoD software acquisition: Status of and challenges related to reform efforts* (GAO-21-105298). https://www.gao.gov/assets/gao-21-105298.pdf

Madachy, R. (2018). *Early phase cost models for agile software processes in the U.S. DoD*. Naval Postgraduate School. https://calhoun.nps.edu/handle/10945/64421

Munroe, R. (2007, December 12). *Nerd sniping.* xkcd. https://xkcd.com/356/

Platform One. (2020). *Platform One intro (extended)* [Video]. https://youtu.be/HFTDlLIlyeA?t=32

Putnam-Majarian, T., & Staiger, J. A., Jr. (2019). Defense trend lines: Leveraging the power of historical data. In A. M. Lungu (Ed.), *Development Research Series. QSM Software Almanac* (5th ed., pp. 31–36). Quantitative Software Management, Inc. https://www.qsm.com/sites/default/files/qsm/2019 QSM Almanac.pdf

Schmidt, E., Work, R., Catz, S., Horvitz, E., Chien, S., Jassy, A., Clyburn, M., Darby, C., Ford, K., Griffiths, J.-M., Louie, G., Mark, W., Matheny, J., McFarland, K., & Moore, A. (2021, March). *Final report: National Security Commission on artificial intelligence.* National Security Commission on Artificial Intelligence. https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf

Serbu, J. (2022, February 7). Pentagon has plan to fix its software development woes. *Federal News Network*. https://federalnewsnetwork.com/defense-news/2022/02/pentagon-has-plan-to-fix-its-software-development-woes/

Automated Data for DevSecOps Programs

William Richard Nichols—is a senior researcher at the SEI with a PhD in Physics. His experience includes data acquisition, scientific software development, nuclear engineering, and software project management. While at the SEI he has coached software project management and published research on using software metrics for project and program management. [wrn@sei.cmu.edu]

Hasan Yasar—is the Technical Director of the Continuous Deployment of Capability group in the SSD Division of the Software Engineering Institute. Hasan leads an engineering group to enable, accelerate, and assure transformation at the speed of relevance by leveraging DevSecOps, Agile, Lean AI/ML, and other emerging technologies to create a Smart Software Platform/Pipeline. Hasan has more than 25 years' experience as a senior security engineer, software engineer, software architect, and manager in all phases of secure software development and information modeling processes. [hyasar@sei.cmu.edu]

Luiz Antunes—is a member of the technical staff at the SEI. With a background in Distributed Computing, Information Architecture & Visualization and HCI, he is responsible for the creation of artifacts that generate visual analytics from large data sets. Prior to joining the SEI, he developed software for the utilities sector in South America and later acted as one of the architects of the CoMotion data visualization and collaboration suite at General Dynamics. He holds a BS in Civil Engineering and a MSc in Hydraulic Engineering from the State University of Campinas, Brazil. [lantunes@sei.cmu.edu]

Christopher L. Miller—is a senior engineer at SEI. [clmiller@sei.cmu.edu]

Robert McCarthy— is at Member of the Technical Staff at SEI supporting DevSecOps engineering. [remccarthy@sei.cmu.edu]

Abstract

Automation in DevSecOps (DSO) transforms the practice of building, deploying, and managing software intensive programs. Although this automation supports continuous delivery and rapid builds, the persistent manual collection of information delays (by weeks) the release of program status metrics and the decisions they are intended to inform. Emerging DSO metrics (e.g., deployment rates, lead times) provide insight into how software development is progressing but fall short of replacing program control metrics for assessing progress (e.g., burn rates against spend targets, integration capability target dates, and schedule for the minimum viable capability release). By instrumenting the (potentially interacting) DSO pipelines and supporting environments, the continuous measurement of status, identification of emerging risks, and probabilistic projections are possible and practical. In this paper, we discuss our research on the information modeling, measurement, metrics, and indicators necessary to establish a continuous program control capability that can keep pace with DSO management needs. We discuss the importance of interactive visualization dashboards for addressing program information needs. We also identify and address the gaps and barriers in the current state of the practice. Finally, we recommend future research needs based on our initial findings.

Introduction

We undertook this research because program management in the DoD face challenges measuring program performance and conducting effective oversight of continuous integration/continuous delivery (CI/CD). Closing this gap should enable adoption of modern practices. To realize the benefits of CI/CD, we investigate how to collect and use metrics from the modern development pipelines to support cost and schedule prediction models derived from that data. Our research project, therefore, examines how to exploit the automation within the DevSecOps (DSO) environment to benefit program management.

Software acquisition increasingly involves software development using CI/CD, as described in the Defense Acquisition University (DAU) training, *Software Acquisition Pathway (SAP) Interim Policy and Procedures*. This SAP training is available to "facilitate rapid and iterative delivery of software capability to the user" (CSO DoD, 2021b) and empower program managers (PMs; Brady & Rice, 2020). However, acquisition program management professionals struggle to keep pace with continuous delivery because it does not come with continuous data or continuous estimation models. Continuous delivery can produce working software not only at the end of sprints, but also daily or even multiple times per day. To make commitments, changes, or program interventions, the program office needs up-to-date information on capability readiness, costs, and progress rates. However, the delivery of relevant data reports can take weeks or months. The actual CI/CD progress is thus constantly ahead of effective program control.

There are many challenges to managing CI/CD intensive programs including the increasing complexity of software-enabled systems, hardware-in-the-loop testing and simulation, and inclusion of COTS and/or open-source components. Within this context, acquisition is adapting by using the SAP. DoD policy is clear; IT expects PMs to use metrics for planning, control, and oversight. As stated in SAP policy, "The PM shall identify, collect, and use management, cost, schedule, and performance metrics to enable effective program execution by the PM and other stakeholders. **Metrics collection should leverage automated tools to the maximum extent practicable**"(emphasis added; "Under Secretary of Defense" DoD, 2020).The specific list of minimum requirements includes process efficiency, software quality, software development progress, cost, and capability delivery (e.g., value delivered).

Background

After conducting an extensive literature review, we found that the peer reviewed literature is devoid of studies about automated data collection for CI/CD (Prates et al., 2019). Although non-peer reviewed literature exists, *it either* addresses operational issues rather than PM issues, *or it* is limited to a narrow research topic rather than DoD *programmatic* needs (Vassallo et al., 2019). Moreover, little attention has been given to managing multiple interacting pipelines, each with a distinct technical stack, personnel, and rates.

Several sources—Practical Software Measurement Group (PSM), National Defense Industrial Association (NDIA), and International Councils on Systems Engineering (INCOSE; Jones et al., 2020b), and the DoD (DoD, 2019)—recommend metrics for Agile and CI, but none connect to automated collection or have metrics that have been rigorously validated. The situation is similar for DSO with regard to the DevOps Research and Assessment (DORA) metrics (Forsgren & Humble, 2015; Forsgren et al., 2018). The Defense Innovation Board (DIB) explicitly identified this gap, "In the beginning stages of the DoD's transformation to DSO methods, the development and operations community will need to work closely with the cost community to derive new ways of predicting how fast capability can be achieved. For example, estimating how many teams' worth of effort will be needed to invest in a given period of time to get the functionality needed. [...] New parameters are needed, and more will be discovered and evolve over time" (McQuade et al., 2019).

By replacing practices that, in the past, have been labor intensive and prone to error, DSO enables CI/CD. CI is the automated process that developers use to integrate code and then build, test, validate, and deploy new applications. The automation that makes these DSO practices possible, in turn, spawns a large amount of data as a byproduct. Making this data available enables stakeholders to assess the health of a project, including its development performance, operational performance, whether it is sufficiently secure, and how frequently upgrades are being delivered.

DSO, Pipelines, and Automation

To implement the automated continuous estimation of software-intensive systems, you must first define what is being measured. The specific measurements depend on the decisions to be made. In this paper, we focus on the decisions made by program managers and the development pipelines as the object of measure.

DSO is a software engineering culture and practice that unifies software development (Dev), security (Sec) and operations (Ops) personnel and their practices. The essential concepts of DSO comprise automating, monitoring, and applying security to all activities of software development. These activities include feature planning, bug fixing, feature development, application and support infrastructure builds ant testing, and releasing new software—whether that involves maintaining operational software that supports a user base, or monitoring operational systems for performance and security-related events (CSO DoD, 2021b).

DSO consists of a set of principles and practices that enable better communication and collaboration among relevant stakeholders for specifying, developing, and operating software and systems products and services and making continuous improvements to all aspects of the life cycle (IEEE, 2021).

A DSO pipeline consists of a chain of processing elements arranged so that the output of each element is the input of the next, much like a physical pipeline. The analogy to a physical pipeline is a weak connection (i.e., there is no requirement for ordered processing or tight coupling). In fact, many DSO pipeline elements use asynchronous messaging and decoupled processes (e.g., GitOps). Often the term *pipeline* is used to describe a set of processes that tie together and eventually produce a software artifact. Sometimes this output is then used as an input into a different, possibly distinct, pipeline or pipeline instance (CIO DoD, 2019; CSO DoD, 2021).

Automation of the DSO pipeline provides an unprecedented opportunity to collect software development data from the engineering tool suite without burdening the software development staff with providing performance metrics, thereby distracting effort from their development work. Eliminating manual data-collection activities not only reduces the effort associated with performing these tasks, it also reduces the opportunity to inject bias into the data. Automated data collection also provides a continuous data collection and storage capability that can revolutionize the frequency and fidelity of software estimation.

Programmatic Needs

Program management is usually defined as managing a group of related projects using specific management techniques, knowledge, and skills. PMs must work with senior leaders and stakeholders across multiple departments and teams. Their decisions are likely to be strategic and connected to the financial calendar. Their responsibilities include coordinating resources and outputs across teams rather than within teams.

PM responsibilities include strategy, finance, and communication. Their overarching purpose is to guide their program to successful outcomes. Specific responsibilities include the following (Zein, 2010):

- Manage the program’s budget.
- Establish high-level performance objectives.
- Manage a strategy, and guide investment decisions.
- Define the program governance (i.e., controls).
- Plan, monitor, and control the overall program.
- Manage risks and issues and implement corrective measures.

- Coordinate the projects and their interdependencies.
- Manage and use resources across projects.
- Manage communication with stakeholders.
- Align the deliverables (i.e., outputs) to the program's outcome with the aid of the business change manager.
- Manage daily program operations throughout the program life cycle.

The PM needs information to provide adequate resources, negotiate commitments, and otherwise satisfy stakeholder needs. The status of any project reflects not only the status of its own code, but also how its dependencies affect it. These needs include, but are not limited to the following:

- baseline and benchmark performance
- product completion and cost rates with probabilistic cost/schedule projections
- a master plan, master schedule, and lead times
- when work begins and is completed
- which queues can be bypassed
- resource needs and resource utilization assuming nominal conditions and "what-if" scenarios

Prior Work/State of the Practice

Prior work has identified numerous candidate measures and opportunities in the DSO pipeline (DoD, 2019; Jones et al., 2020b; McQuade et al., 2019) at all stages of development, including feature request, requirements development, architecture, design, development, test, delivery, and operations. Automatically collecting data generated by the tools used during these stages can provide information on product size, effort, defects, rework, and durations—often on a feature, story, and/or component level of granularity. A key challenge is creating new features, in the machine-learning sense, from the raw data to deduce status and improve predictive power. Nonetheless, peer-reviewed research is severely limited for programmatic metrics. In this section, we summarize the current literature on DSO metrics research and practice.

Academic Research

The research community has only recently begun to study measurement in DSO. A multivocal literature review by Prates found limited prior academic research about DSO metrics (Prates et al., 2019). Moreover, the metrics that Prates identified focused on security and quality (e.g., defect burn rate, critical risk profiling, defect density, top vulnerability types, number of adversaries per application, adversary return rate, point of risk per device). Prates' summary noted "It was very hard to find information regarding metrics associated with DSO in academic literature." Primarily, the metrics identified were security related rather than programmatic. In a 2020 paper, Mallouli focused on cybersecurity rather than programmatic issues (Mallouli et al., 2020). A more general contribution from Mallouli included a metrics-driven DSO architecture that includes measuring tools, a core platform, a database, and analysis tools. Mallouli's architecture diagram aligns with our vision of general-purpose needs for DSO measurement.

The Government and DSO

One defining characteristic in the DoD is that the environments in which systems operate are highly regulated. Because of this, agencies are not free to simply adopt strategies and frameworks from industry environments. The Software Engineering Institute (SEI) has written guidance that describes special conditions found in these environments, difficulties generated by them, and possible solutions to make DSO practices work (Morales et al., 2020).

One of the biggest pushes to Agile, DevOps, and DSO started after appointing a Chief Software Officer at the USAF. The DoD has since undertaken an effort towards the internal standardization of a platform with artifacts and processes that may be used across departments and agencies. While not a one-size-fits-all solution, this initiative has promoted the DSO mindset across multiple programs that now is the right time to implement DSO practices. To support these initiatives, the DoD prepared guidance on how to adopt DSO practices and provided ideas about teams and personnel organization, cost, and levels of effort. Most of these departments and agencies follow the guidance in these DoD documents:

- The **DoD Enterprise DevSecOps Playbook** provides detailed coverage of all aspects of the design, development, and operation of systems under the DSO lens. The topics covered include shifting a program culture towards DSO, assembling a software factory (SWF), implementing DSO pipelines in an SWF, capturing basic metrics monitor progress, orchestrating frameworks, and securing a system and its infrastructure (CSO DoD, 2021a).
- The **DoD Enterprise DevSecOps Reference Design** preceded the playbook mentioned above. This document provides technical implementation details such as selecting containers versus virtual machines, using a DoD centralized-artifact repository, and organizing a DSO pipeline and its environment (CSO DoD, 2021b).

Industry

Industry tends to be on the “bleeding edge” of technology and is always adopting practices that can provide a competitive advantage to its organizations. Much of the guidance initially adopted by industry comes from documents published by consortiums of organizations that have a solid track record in implementing DSO and software factories that apply advanced concepts to be more competitive and secure (e.g., Netflix Chaos Monkey and Amazon fast-turnaround live-release deployments).

DSO Measurement and Metrics

Measurement of DSO draws from both traditional Agile (Kupiainen et al., 2015), Lean (Poppendiek & Poppendieck, 2013; Staron et al., 2012), and flow (Vacanti, 2015) metrics. Measurement objectives include tracking project progress, increasing visibility into complex aspects of development, providing adequate resources, balancing workloads, understanding and improving quality, ensuring adequate testing, and verifying readiness for release (Kupiainen et al., 2015). This section includes descriptions of adaptations of metrics common in DSO.

Using analysis surveys completed by DevOps subject matter experts (SMEs) the DORA (Forsgren et al., 2018) identified four key metrics associated with software development and delivery performance. Two metrics relate to tempo, two to stability, and one to reliability.

- **Deployment frequency** is the frequency of an organization’s successful releases. Because different organizations define *release* differently, deployment frequency might measure how frequently code is deployed to preproduction staging, to production, or to end consumers. Higher frequency is considered better.
- **Mean lead time from commit to deploy** is the mean lead time for change or the average time required for a commit to reach production. Short mean lead times enable engineering and management to determine that the post-code production process is healthy and likely could support a sudden increase of requests. This metric, like deployment frequency, is a measure of software delivery speed.

- **Mean time to recover (MTTR)**[47], aka mean time to restore, is the average duration in time required to restore service after an unanticipated issue or outage. Short recovery times are enabled by rigorous monitoring, full configuration control, infrastructure as code, and automation that enables a prompt roll back to a stable system. Shorter outage durations and recovery times are better.
- **Change failure rate** is a percentage that measures the frequency at which changes to the production system result in a problem including rollbacks, patches, and failed deployments. A lower change failure rate is better and indicates the production process is effective. Higher rates indicate that developer time is spent on rework rather than new value.

The General Services Administration (GSA) provides a larger set of metrics for measuring the success of implementing DSO (GSA, 2021). These high-value metrics include deployment frequency, change in lead time, change in volume, change in failure rate, mean time to restore, availability, customer issue volume, customer issue resolution time, time to value, time to authorization to operate (ATO), and time to patch vulnerabilities.

The PSM issued three framework documents for measuring continuous iterative development (CID). *PSM CID Measurement Framework Part 1* describes the concepts and definitions (Jones et al., 2020b), *PSM CID Measurement Framework Part 2* addresses measurement specifications and enterprise measurement (Jones et al., 2020a), and *PSM CID Measurement Framework Part 3* addresses technical debt (Jones et al., 2021).

Our Research

In this section, we describe our research objectives, approach, workflow, and early results.

Objectives

To demonstrate the feasibility of automated continuous measurement and estimation we simulated a software project using synthetic data and a prototype instrumented DSO pipeline (Abdel-Hamid et al., 1991; Raffo, 2004). In the demonstration, we focused on a subset of DoD PM information needs, leaving a more comprehensive effort for future work. We also focused on projections for satisfying the requirements for coordination dates such as the minimum viable product (MVP) or minimum viable capability release (MVCR). To validate our work, we used quarterly advisory review panel sessions (QuARPs) involving DoD PMs and other SMEs.

Our long-term research goal is to improve the support of PM decision making. The short-term objective was to explore the subject for gaps, needs, and research opportunities. Thus, a successful project would lead to more focused follow-on work. With all these objectives in mind, the research team posed a number of related questions to explore, including the following:

- What information gaps do DoD PMs have with DevOps-related projects?
- What program information is needed for prediction and actionable decisions in this environment?
- What data supports answering those questions?
- What data can we gather to support real-time reports and analysis?
- How should the data be joined, transformed, and labeled to retain the context?

---

[47] https://en.wikipedia.org/wiki/Mean_time_to_recovery

- What algorithms should we use to develop models and indicators?
- How should we present indicators to decision makers?

The above questions can be binned into three ideas that guide our research:

- What a PM needs to know from a software CI/CD pipeline
- How progress against goals can be measured using this information
- How the information should be presented

Approach

Our approach to this research included the following steps:

- Identify SMEs for consultation and review.
- Select key program management scenarios.
- Construct a prototype pipelines
- Hypothesize performance indicators.
- Prototype pipelines of and collection of data
- Predict performance using synthetic data.
- Validate dashboards with SME

Because we entered this research with assumptions, our SME proved invaluable challenging, validating, and elaborating on use cases and workarounds. They guided us to focus on percent complete, predictions of capability delivery dates with the status quo, and predictions of capability delivery dates with program interventions.

We selected SMEs who had significant responsibilities in the DoD and defense industrial base in areas such as program management, DSO consulting, and government policy. Although selecting SMEs risked introducing bias, the benefit was a small group with whom we could engage in deeper discussions. We constructed a demonstration DSO pipeline with instrumentation points for prototyping data collection and storage. We reviewed this pipeline with our SMEs to verify that it addressed their concerns.

To make decisions, a decision maker must have information about the scenarios. We borrowed indicators typical of earned value (*Department of Defense Earned Value Management Interpretation Guide*, 2018) and earned schedule (Lipke, 2003) management and validated these indicators with the SMEs.

Based on information needs, we explored the prototype pipelines and other data sources. This helped us identify data sources and reason how to collect the data with the sufficient context to construct the indicators.

Using actual data was impractical because of the limited time available for completing our work. Instead, we generated synthetic data, which was suitable for our purposes and provided additional benefits. The purpose of the simulation was to demonstrate that our data could be stored and that our data storage models would be suitable for producing the desired indicator.

We began with a hypothetical project. We separated the work items into capabilities, features, and stories to develop a reference roadmap and work breakdown structure (WBS) and distributed work among two hypothetical teams. Next, we added an artificial estimate for direct effort to each story. We approximated duration as proportional to effort and parameterized the variation in the actual effort required. We used a nominal team load and effort calendar to map the beginning and end of development for each component to an initial estimated plan schedule and an actual (simulated) schedule.

We simulated the flow of stories through our pipelines to model data collection and migrated the data into a database. We then extracted data from the database to build the indicators. We computed the percent complete based on estimated costs and estimated costs of complete work. The results were displayed as an earned schedule. We computed projected scenarios using the *a priori* effort variation and Monte Carlo to estimate a range of completion dates. We then demonstrated the simulations and resulting indicators to our SMEs for their review.

Workflow

The collection of data throughout the design, development, and operation a system provides the people involved in these processes with situational awareness and actionable information.

Information from processes and tooling can be captured from the early stages of planning, throughout the execution of the system, and finally from the environment in which the system operates (in a post-deployment scenario). Figure 1 displays the different phases of the system life cycle and suggests types of data that could be collected along the way.

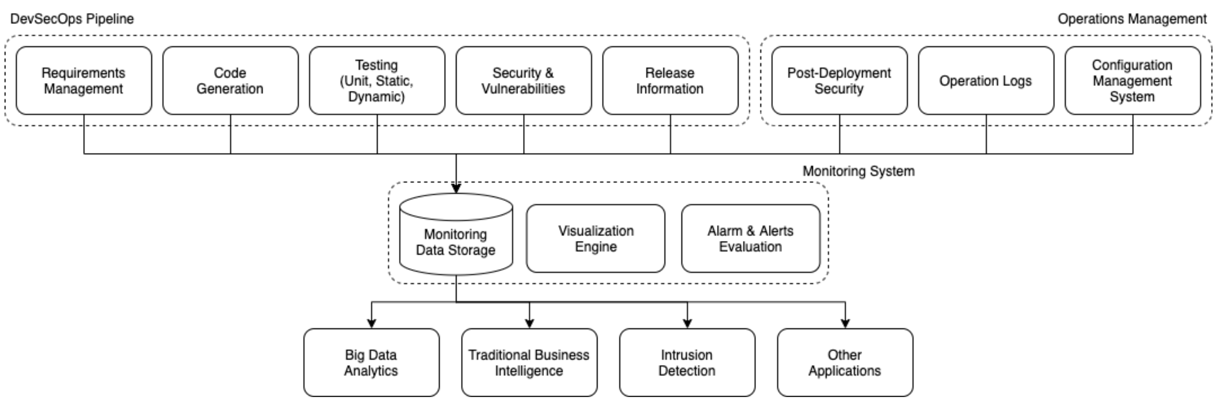


Figure 19. DSO Pipeline and Data Storage

Recorded requirements can be monitored from the planning stages of the system through the development phases to ensure that features are implemented according to the original plan. Because specifications may change along the way, changes to these requirements can and should be incorporated to tell a complete story and indicate the reasons why modifications in the implementation are necessary.

As the design and architecture phases begin, these requirements take shape and—based on architectural principles—are turned into a “skeleton” that will guide high-level feature generation. Code is then generated to implement the features proposed by the architecture and is then refined by epics and stories.

At the same time, artifacts enter a version control system and start flowing through a CI framework that allows data to be captured, such as code style, quality, and security. This data can be inspected and discovered by a linting and a static analysis process.

In the CI framework, the system is built and tested by a dynamic analysis process that evaluates the quality and security of the built system and its dependencies, which were detected in the build process. The data collected from this phase is extremely valuable to the teams involved in feature implementation because it guides them through fixing issues and minimizing risk. This data is also useful for teams managing resources and following the cost and schedule because it (1) informs them about the efficacy of the development team during implementation and (2) helps forecast estimated dates and the overall cost of completion.

As development teams release system versions and move them to staging environments—and finally production—all the data about post-deployment issues and system utilization can also be captured to inform operation teams about resource utilization and system growth.

There is so much data that anyone who monitors it can feel overwhelmed by its volume and what it covers. That is why it is important to introduce mechanisms that reduce mental analysis, make good use of human cognitive capabilities, and allow people to form faster insights. Stakeholders need information that helps them answer questions about system planning, development, and operation. Understanding their needs enables us to create conditions for better sustainment, faster problem solving, and increased security.

In this study, our teams analyzed data captured in many of the different phases described above, for both real and simulated projects. Our research team wanted to answer questions that might have a significant impact in the development and operation of the system, and they chose metrics based on their analysis. Once we defined metrics, our developers introduced means to capture and store the data supporting to those metrics and generated visualizations that could make the data easier to understand. Based on stakeholders needs, these visualizations we aggregated the metrics into dashboards that now provide full transparency into the development and operation of the system.

The fact that we are paying attention to not one pipeline, but an association of pipelines introduces complexity to capturing and organizing data while understanding (1) the origin of the data and process sequencing and (2) classifying and separating information at presentation time. Such a significant amount of information can be overwhelming and become extremely misleading if the design of the dashboards does not provide enough situational awareness to those consuming the information.

CONOPS

The overall goal of this research is to understand the behavior of the processes in the system development life cycle (SDLC) by capturing measurements that provide situational awareness of the efficiency of the different parts of the framework and system. However, the complexity introduced by the interactions across multiple frameworks and pipelines can add substantially to how much data is monitored in the different parts of the environment. Teams must be careful when introducing instrumentation to ensure that it provides an accurate view of different paths and considers the right timing for measuring signals.

Figure 2 is a concept of operations (CONOPS) that illustrates the different phases of the overall process, including planning, implementation, and operation. All phases provide valuable information for monitoring and thus creating an accurate situational awareness model for stakeholders.

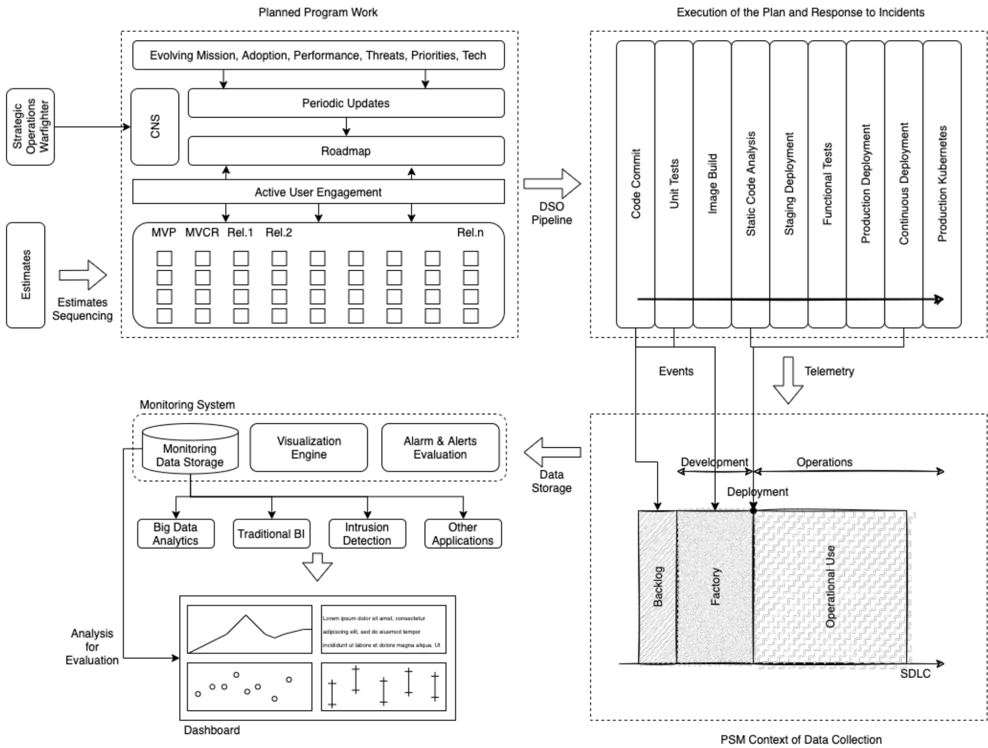


Figure 20. Measurement Collection CONOPS

Before teams implement the mechanisms described in the Workflow section, they must understand what the organization sponsoring the development of the system expects and what plan has been generated for this process. The initial plan contains estimates of complexity for different system modules as well as forecasts of the schedule and costs involved in each phase. Part of the reporting is generated by comparing this plan with its execution. Because the SDLC is using Agile methodologies, changes to requirements are always welcome and must feed back into estimates at the end of each iteration, making this whole process more dynamic.

Approaching the production of the system, a DSO pipeline—or group of pipelines—provides access to a large amount of information that can be captured directly from the tooling used in the pipelines. Every interaction with the framework allows information for monitoring to become available, such as code style and quality, secure coding practices, results of unit tests, static code analysis, dynamic application analysis, functional testing results, container security testing results, and staging/production environment analysis results.

All of this information should be properly captured and made available to stakeholders through visualizations and dashboards—or alerts and alarms for critical and more urgent events. This approach makes it possible to introduce adjustments to system construction and operation estimates and initiate corrective actions that will generate a positive impact in time to develop or correct system issues.

Early Results

We presented common PM scenarios and questions that might require measurement to support PM evaluations or decisions. Our intent was to prioritize programmatic needs for immediate focus rather than identify all programmatic needs. We recorded the results of these discussions, categorized the questions as "Status and Projections" and "What If," and summarized the results in Table 1.

Table 13. Program Scenarios

Scenario 1: Status and Projections	Scenario 2: What If?
Will we meet the schedule commitment?	**Can we accept a change?**
Where are we now?	**What if we reduce the scope?**
What is our completion rate?	**What if we add resources?**
• How much actual effort was applied?	• What is the required effort?
• Which items are complete?	• How will our completion rate change?
• Which items remain for each capability?	• How are capability commitments affected?
• What is the percentage complete overall and per capability?	
When will we finish the current work?	**If we add effort, how long will it take?**
• What is the projection for completing the project, including schedule and cost estimates?	• What is the new projection for completing the project?
• What is the projection for when each capability will be fully realized, including schedule and cost estimates?	• What is the new projection for when each capability will be fully realize?
• What is the confidence range of current estimates?	• What is the confidence range?
• What are the completion rates and the amount of estimation bias?	
• What are the rework rates?	

Scenario 1 (Status and Projections) focused on the project's status. Status requires understanding the overall body of work, the specific work complete, and the planned and actual cost and schedule for that work. Of specific interest in CI/CD development is the percentage complete overall and for specific capabilities. After the advisory review panel reviewed the primary scenario, it also wanted projections for schedule and cost at completion for each capability and sets of capabilities. In addition, it also requested credible ranges of cost and schedule. These additions were considered important for making commitments and planning resources.

Scenario 2 (What If?) focused on making decisions about program interventions. Typical interventions include changing priorities, increasing, or decreasing scope, and shifting resources. For each of these interventions, the panel wanted a credible range of estimates before and after the intervention.

Although these are typical PM concerns, having timely information has been problematic because of the following:

1. Information was scattered across different systems.
2. Information across the systems, even if available, was not easily joined.
3. The measurements were seldom at the level needed to answer the necessary questions.

For example, if stories recorded during a sprint could be traced to different capabilities, then the following problems could occur:

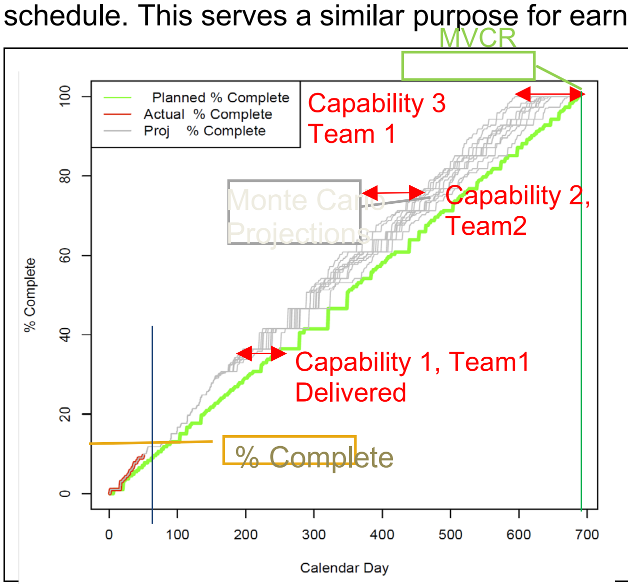
- External mappings would be needed to determine capability completion.
- Effort variances could not be distinguished among capabilities or types of work.
- Variation information would be limited to the sprint rather than to the story level.
- Projections would require detailed knowledge of the planned work order.
- Capability work could be spread across different teams.

The continuous measurement of start and completion times for each story helps resolve some of these problems, but that measurement still relies on fitting information together from the WBS, the master plan, and master schedule. Successful PMs described resolving some of these issues using pivot tables. This is a manual solution to the data join problem, but it does not fully address the unit of measure or analysis problems.

Indicator Displays

We provide prototype indicator for status in Figure 3. These indicator uses data from a simulated project. This indicator shows the plan and projected delivery for each three capabilities developed by two different teams. We structured a representative project into capabilities, features, and stories. We estimated work and sequenced it for execution. We parameterized work package duration with lognormal distribution for actual duration uncertainty and a small underestimation bias was introduced. We separately measured rates and variances for each of the teams. The Planned line represents the rate of progress of the sequential execution of the work packages assuming estimated effort was both available required. The Actual line represents a Monte Carlo simulation though 10% of the estimated duration. The Projection line measures the estimation bias and variation, then it applies the empirical bias and variation to the remainder of the work packages. A number of Monte Carlo simulations then show a range of probable dates, enabling a 90% likelihood estimate. The significance of this simulation is that data is collected automatically from tools using the events defined using Figure 4, Example Sequence Diagram Between Commit and Deploy.

Percentage complete against scheduled cost indicates an earned value. A horizontal line from the work complete to the plan provides a visual representation of days ahead or behind schedule. This serves a similar purpose for earned schedule (Lipke, 2003).



We next presented the SME with a graph, showing the effects of moving half the work to a second team and rebalancing work as needed. This graph represents one of many possible program interventions. Although we recognize that this is an oversimplification, the presentation was adequate for the purpose of obtaining SME validation for the requirement. The SMEs agreed that a similar graphic to compare the current likely outcomes with a probabilistic range of completion dates after an intervention is needed.

Other interventions not included in this paper included adding or removing capabilities and shifting commitment dates. It is a straightforward matter to indicate the completion of specific capabilities along the timeline.

Supporting Metrics

For the purposes of these simulations, we made simplifying assumptions. At this stage, our objectives were to validate the displays with the SMEs and verify the data-collection approach. The following are the simplifying assumptions we made:

- Estimation bias from completed items continues (i.e., the average completion rates will continue to follow the historic trend).
- The estimation error will distribute lognormally.
- Applied effort (cost) is accurately recorded and projected.
- Effort in labor days has been entered for each capability and feature.
- The relative size of stories has been converted into effort days.
- Story effort equals the development duration in labor days.
- A story is worked by only a single developer.
- The stories are worked sequentially in a batch size that does not exceed the number of developers.

Metrics supporting these indicators include the following:

- percent complete (i.e., the estimated cost of all capabilities/estimated cost of capabilities complete)
- completion rates
- schedule projections (i.e., Monte Carlo projected completion date for each sequenced story)

The measures include the following:

- capability, feature, and story estimates in labor days
- story start date
- story completion date from deployment
- story effort (i.e., the development duration in labor days)

Tool Sequence and Data Collection

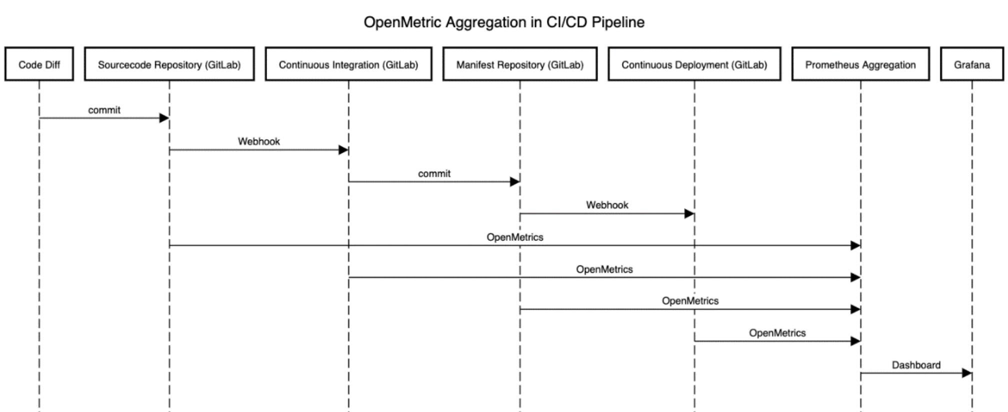


Figure 22. Example Sequence Diagram Between Commit and Deploy

Almost all CI/CD tools offer some sort of collection endpoint, such as an API. These endpoints offer structured data in predefined formats that allow for the collection of metrics regarding builds, health, load, and frequency of use (among others).

Clients are responsible for generating their own metric endpoints for an aggregator to consume. However, the format the tools use to output their data must be standardized. OpenMetrics offers a standard format that displays this data for aggregation engines to consume. This format ensures that metrics are newline separated, with their key:value space separated. This simple format also allows tagging metrics with any number of labels and adding context to each metric where appropriate.

A data aggregation engine like Prometheus can be configured to point to these endpoints for data collection and point to itself to collect data about its metric outputs. Prometheus servers can also be distributed to have a central collection point in the context of several pipelines, which requires aggregated statistics as described in Pipelines of Pipelines (PoPs). (Prometheus can be installed either as a standalone server or within a Kubernetes cluster via Helm or as an Operator.)

DoD customers and the government may leverage techniques, such as Federation, to retrieve and manage aggregate statistics about various vendor pipelines as development takes place.[48]

Once data is flowing into a metric aggregation engine such as Prometheus, tools like Kibana or Grafana can be used to further visualize that data. These visualization tools can be used to create custom dashboards for keeping operators informed in real time about changes in pipelines across a number of DoD projects.

Discussion, Open Issues, Next Steps

Lessons Learned

During this prototyping, we identified several issues that must be overcome to achieve the desired ability to measure schedule and cost progress.

Capability-Based Work Breakdown Structure

The first issue is obvious: The product roadmap needs to be sufficiently developed to estimate the entire scope of work contained in the capability. We are aware that a project's scope will often change, but a nominal scoping and initial estimate are a minimum requirement.. It is critical that traceability of the work package (feature, story, or task) to the capability be maintained throughout the project. It is not, however, required that all stories related to a capability be done to the exclusion of other work or that they be done in a specific order. Nonetheless, the sequencing of features and stories define the up-to-date master plan, which determines the master schedule. Variances from that order must be recognized, as do changes to the work scope. A capability is complete when the last task associated with that capability is released. Although this seems straightforward, rework complicates its use in practice.

Reworking stories or adding defect fix stories confounds this approach. We recommend not counting stories as complete until they are thoroughly tested and released. Defect fixes should be included as separate stories that do not count toward the earned schedule, but that do consume resources. This can be accomplished by adding defect fixes to WBS elements that do not contribute to the earned schedule but that do require flow through the system. This also has the effect of adding cost and schedule, but not adding progress to percent complete.

Connecting the Stories to the WBS

The traceability of stories to the WBS is not directly supported by existing tools. Although workflow is often managed by Jira, some instances use GitLab or other tools. Typically, these workflow management tools do not link directly to the roadmap or WBS. The mapping can be

---

[48] Learn more about Federation on the Prometheus website (https://prometheus.io/docs/prometheus/latest/federation/).

overcome with the careful use of labels. However, labeling requires consistency and is error prone. An alternative is to maintain a separate mapping between WBS elements and their representation in the workflow tool. As long as the mapping is maintained, the story flow can be traced through the DSO tool chain.

The instrumentation of a pipeline versus a pipeline instance poses another problem. Several arbitrary ways exist to organize similar DSO tool chains. Different tools can provide similar functionality but have different interfaces. Some tools might have different orders of execution. The instance must be described sufficiently so that the actual progress of a story is known, and that the data can later be used with its context. In principle, there should be ways for a toolchain to describe itself. Nonetheless, an automated tool chain should be repeatable and stable. For this reason, we characterized a pipeline instance by its activity and by which tool performed or was used in performing that activity. To effectively use automated data collection, events from the example sequence diagram in Figure 4 for capability or analysis must trace the work package to the specific capability and feature.

The biggest gap in data collection is the start of work. Once the story achieves code completion, the automation accurately tracks progress, including rework. However, designating the start of work can be problematic. Currently, we rely on an entry to the workflow management tool as the start and entry to the DSO deployment tool for completion.

Data Warehouse vs. Data Lake

We considered using both a data lake and a data warehouse in our design. The primary difference is that the data lake follows an extract, load, transform model, while the data warehouse follows an extract, transform, load model. Both begin by extracting data from the system. However, while the data lake loads the data into storage, the data warehouse transforms the data by performing logical joins and adding related contextual information prior to loading it into storage.

Using the data warehouse, data can be retrieved after it is loaded and instantly be used to build pre-defined indicators. The warehouse is efficient because the transformation is applied only once, and the structure can be tuned to support the desired indicators. The drawback is that support of other indicators or uses can be inefficient and cumbersome. Nonetheless, designing the warehouse requires forethought into the required context that will be needed. If this context was not stored or is not accessible, the indicator may not be possible to build.

The data lake, on the other hand, delays transformation until the data is used. This is inefficient for repeated operations because the transformation must be applied every time the data is used. However, late transformation provides more flexibility to use data for other purposes and new indicators. In practice, a workable approach is to stage the data in a data lake and immediately extract and transform into a warehouse. Although this, is inefficient for storage, it supports both needs for repeated use and research.

Opportunities for Further Research

In this research, we identified gaps where the state of the practice does not fully support the needs of defense acquisition. Although some of the gaps apply predominantly to DoD needs, their solution has more general application for all organizations. Additional gaps remain; for example, a recent DoD memorandum that addressed continuous authority to operate (cATO; McKeown, 2021) states the following:

- Service providers will continuously monitor and assess all of the security controls within the information system's security baseline, including common controls.
- Automated monitoring should be as near real time as feasible.

- For cATO, all security controls will need to be fed into a system-level dashboard view, providing a real time and robust mechanism for AOs to view the environment.

Automation of data collection from DSO pipelines promises to address this and other information needs. We foresee future research that includes the following:

- modeling parametric cost estimation as the program evolves
- extension to software factories and multiple interacting pipelines
- inclusion of quality, rework, and technical debt in management goals
- modeling cybersecurity authorization and risk

Summary

In our review of DSO metrics practice, we found limited integration of DSO measurement into program management decisions. Identifying measures, validating measures, and providing a supporting infrastructure remain largely unexplored.

This research focuses on improving program management decision making by improving the fidelity and frequency of program performance metrics and indicators, including information needs, what to measure, and how to display the information.

SMEs provided the research team with key program management scenarios to focus our research. We created prototype pipelines to provide a frame of reference for generating candidate indicators of program performance. Using synthetic data, we simulated software development activity. We used the data to build indicators that we validated with SMEs. The overall workflow that we created and captured provides a unique conceptual view of how data can be extracted, stored, and reported from an Agile and DSO pipeline.

A year into this research project, we have several lessons that are worth sharing:

- **Adopt a capability based WBS**. The most fundamental information an organization's leadership wants to know is, "When will it be done?" In a DSO environment, *done* is measured by delivered capabilities; therefore, aligning a WBS to capabilities is an essential first step.
- **Connect engineering artifacts (e.g., stories) to the WBS and associated work packages**. When performance indicators reveal failure to meet the plan, PMs then ask the question "Why not?" To drill down into the data and identify the source of the discrepancy, the cost and schedule targets must align to engineering activities, subsystems, or even individual components.
- **Establish a robust analysis capability in conjunction with creating and maintaining a sufficient data storage system**. The types of analyses and robustness of reporting drive the data storage requirements. The data to be collected and stored drives data infrastructure design considerations. The information needs of the organization drive data warehousing and data lake design options.

As this research continues, it will focus on refining and improving the collection, storage, and reporting of project performance data that is most needed program management. We identified important areas, but we did not include them in the scope of this research. These areas pose great challenges and include parametric cost estimation modeling; collection tooling and application programming interface (APIs); quality, rework, and technical debt; and cybersecurity. While each of these is significant in its own way, our research is tackling the challenges associated with integrating software factories and multiple pipelines in the upcoming year.

References

Abdel-Hamid, T., & Madnick, S. E. (1991). *Software project dynamics: An integrated approach*. Prentice-Hall, Inc.

Brady, S., & Rice, C. (2020). *Software acquisition pathway interim policy and procedures training*. https://www.dau.edu/Lists/Events/Attachments/193/SW Acq Policy Training_3.17.20.pdf

CSO DoD. (2021a). *DoD enterprise DevSecOps fundamentals playbook*. https://software.af.mil/wp-content/uploads/2021/05/DoD-Enterprise-DevSecOps-2.0-Playbook.pdf

CSO DoD. (2021b). *DoD enterprise DevSecOps reference design*. https://dodcio.defense.gov/Portals/0/Documents/Library/DevSecOpsReferenceDesign.pdf

DoD. (2019). *Department of Defense agile metrics guide*.

Department of Defense Earned Value Management Interpretation Guide. (2018). https://acqnotes.com/wp-content/uploads/2014/09/DoD-Earned-Value-Management-Interpretation-Guide-Jan-2018.pdf

Forsgren, N., & Humble, J. (2015). DevOps: Profiles in ITSM performance and contributing factors. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2681906

Forsgren, N., Humble, J., & Kim, G. (2018). *Accelerate*. IT Revolution.

GSA, C. (2021). DevSecOps guide. Retrieved September 15, 2021, from https://tech.gsa.gov/guides/dev_sec_ops_guide/

IEEE. (2021). IEEE standard for DevOps: Building reliable and secure systems including application build, package, and deployment. *IEEE Std 2675–2021*. https://doi.org/doi: 10.1109/IEEESTD.2021.9415476

Jones, C. L., Draper, G., Golaz, B., Martin, L., & Janusz, P. (2020a). *Practical software and systems measurement continuous iterative development measurement framework part 3: software assurance and technical debt*.

Jones, C. L., Draper, G., Golaz, B., Martin, L., & Janusz, P. (2020b). *Practical software and systems measurement continuous iterative development measurement framework PSM continuous iterative development measurement framework*. Washington, D.C.

Jones, C. L., Golaz, B., Draper, G., & Janusz, P. (2021). *Practical software and systems measurement continuous iterative development measurement framework part 2: Measurement specifications and enterprise measures*.

Kupiainen, E., Mäntylä, M. V., & Itkonen, J. (2015). Using metrics in Agile and Lean software development—A systematic literature review of industrial studies. *Information and Software Technology*, *62*(1), 143–163. https://doi.org/10.1016/j.infsof.2015.02.005

Lipke, W. H. (2003). Schedule is different. *The Measurable News*, *2*, 31–34. http://www.pmi-cpm.org/members/library/Schedule Is Different.lipke.pdf

Mallouli, W., Cavalli, A. R., & Bagnato, A. (2020). *Metrics-driven DevSecOps, (Icsoft)*, 228–233. https://doi.org/10.5220/0009889602280233

McKeown, D. (2021). *Memorandum for senior pentagon leadership: Continuous authorization to operate (cATO)* (DoD S. I. S. O.). Office of the Secretary of Defense. https://dodcio.defense.gov/Portals/0/Documents/Library/20220204-cATO-memo.PDF

McQuade, J. M., Murray, R. M., Louie, G., Medin, M., Pahlka, J., & Stephens, T. (2019). *Software is never done: Refactoring the acquisition code for competitive advantage* (supporting information).

Morales, J., Turner, R., Miller, S., Place, P., & Shepard, D. J. (2020). *Guide to implementing DevSecOps for a system of systems in highly regulated environments* (No. CMU/SEI-2020-TR-002).

Poppendieck, M., & Poppendieck, T. (2013). *Lean software development: An agile toolkit*. Addison-Wesley.

Prates, L., Faustino, J., Silva, M., & Pereira, R. (2019). DevSecOps metrics. *Lecture Notes in Business Information Processing*, *359*, 77–90. https://doi.org/10.1007/978-3-030-29608-7_7

Raffo, D. M. (2004). Using software process simulation to assess the impact of IV&V activities. In *ICSE 2004* (pp. 197–205). https://doi.org/10.1049/ic:20040459

Staron, M., Meding, W., & Palm, K. (2012). Release readiness indicator for mature agile and lean software development projects. *Lecture Notes in Business Information Processing*, *111 LNBIP*, 93–107. https://doi.org/10.1007/978-3-642-30350-0_7

"Under Secretary of Defense" DoD. (2020). *Software acquisition pathway interim policy and procedures* (US DoD USA002825-19).

Vacanti, D. S. (2015). *Actionable agile metrics for predictability: An introduction*. ActionableAgile Press. https://leanpub.com/actionableagilemetrics

Vassallo, C., Proksch, S., Gall, H. C., & Di Penta, M. (2019). Automated reporting of anti-patterns and decay in continuous integration. *Proceedings—International Conference on Software Engineering*, *2019-May*, 105–115. https://doi.org/10.1109/ICSE.2019.00028

Zein, O. (2010). Roles, responsibilities, and skills in program management. In *2010 PMI Global Congress Proceedings—EMEA*. http://www.pmi.org/learning/library/roles-responsibilities-skills-program-management-6799

New Effort and Schedule Estimation Models for Agile Processes in the U.S. DoD

Raymond Madachy—is a Professor in the Systems Engineering Department at the Naval Postgraduate School. His research interests include system and software cost modeling for total ownership cost; affordability and tradespace analysis; modeling and simulation of systems and software engineering processes, and the DoD acquisition process; integrating systems engineering and software engineering disciplines, and empirical-based research with process simulation. His books include Software Process Dynamics, co-authoring Software Cost Estimation with COCOMO II, Software Cost Estimation Metrics Manual for Defense Systems, and What Every Engineer Should Know about Modeling and Simulation. He is currently writing Systems Engineering Principles for Software Engineers. [rjmadach@nps.edu]

Wilson Rosa—serves as Assistant Director at the Department of Homeland Security Cost Analysis Division. He earned a PhD and an MS in Engineering Management from the George Washington University and a BS in Mechanical Engineering from the University of Puerto Rico. Rosa completed the Senior Managers in Government Executive Leadership Program at the Harvard Kennedy School. [wilson.rosa@hq.dhs.gov]

Bradford K. Clark—is Vice-President of Software Metrics Inc., a Virginia based consulting company. His area of expertise is in software cost and schedule data collection, analysis and modeling. Clark received a PhD in Computer Science in 1997 from the University of Southern California. He co-authored a book with Barry Boehm and others on Software Cost Estimation with COCOMO II. Clark is a former U.S. Navy A-6 Intruder pilot. [brad@software-metrics.com]

Abstract

The DoD's new software acquisition pathway prioritizes speed of delivery, advocating agile software processes. Estimating the cost and schedule of agile software projects is critical at an early phase to establish baseline budgets and to select competitive bidders. The challenge is that common agile sizing measures such as story points and user stories are not practical for early estimation as these are often reported after contract award in DoD.

This study provides a set of parametric effort and schedule estimation models for agile projects using a sizing measure that is available before proposal evaluation based on data from 36 DoD agile projects. The results suggest that initial software requirements, defined as the sum of functions and external interfaces, is an effective sizing measure for early estimation of effort and schedule of agile projects. The models' accuracy improves when application domain groups and peak staff are added as inputs.

Keywords— Agile software processes, Cost estimation, Requirements/Specification, Software acquisition, Time estimation

Introduction

In the United States Department of Defense (DoD), the cost and schedule estimation of agile software development projects is more critical early in the life cycle when limited data is available. These estimates are needed for evaluating contractor cost proposals (Alleman et al., 2003) and to establish initial program budgets and schedules. Accurate estimates (Jorgensen, 2005; Nan & Harter, 2009; Shepperd & Schofield, 1997) help minimize cost overruns and schedule delays (Pendharker et al, 2005; Pillai & Sukumaran, 1997).

Since 2003, more than 1,000 software project data reports (DoD, 2020b) have been collected in the DoD. Of those, less than 5% were identified as agile. The lack of agile software project data has hindered the DoD's ability to implement accurate estimating methods and to articulate whether adopting agile could result in significant savings (Molokken-Ostvold & Jorgensen, 2005). The problem is compounded as agile software processes are increasingly used in the DoD, and acquisition practices must keep pace with these changes.

Studies on agile estimation have either used story points (Bilgaiyan et al., 2016; Choetkiertikul et al., 2018; Choetkiertikul et al. 2019; Owais & Ramakishore, 2016; Nguyen-Cong & Tran-Cao, 2013; Popli & Chauahn, 2015; Usman & Britto, 2016), user stories (Alleman et al., 2003; Saini et al., 2018), function points analysis (Bilgaiyan et al. 2016; Garg & Gupta, 2015; Kang et al., 2010; Nguyen-Cong & Tran-Cao, 2013; Usman & Britto, 2016), use case points (Bilgaiyan et al. 2016; Nguyen-Cong & Tran-Cao, 2013; Usman & Britto, 2016), COSMIC method (Nguyen-Cong & Tran-Cao, 2013), or object-oriented (Alshayeb & Li, 2003) artifacts as primary sizing measures. A second considerably smaller set of studies (Choetkiertikul et al., 2018; Choetkiertikul et al., 2019; Owais & Ramakishore, 2016; Popli & Chauhan, 2014) revealed story points is related to schedule. Although these are widely accepted agile sizing measures, using them at early phases is challenging (Tanveer et al., 2016) as these are typically available later in the life cycle (Choetkiertikul et al., 2019; Jones, 2013; Malik, 2011; Nassif et al, 2013). In the DoD, these sizing measures are provided by developers after contract award (Kaya & Demirors, 2011; Ochodek et al., 2011). Consequently, there is a dire need for early phase cost models (Jorgensten & Grimstad, 2011; Popli & Chauhan, 2013) to help programs get funding before contracting for agile software development projects.

This study contributes to the knowledge base by introducing simple models for estimating effort and duration of agile software development projects at an early phase. An important distinction of this approach is choosing sizing measures (Abraham & Insfran, 2008; DoN, 2010; Malik & Boehm, 2011; Sharma & Kushwaha, 2010, 2012) as model inputs (Ebrahimi, 1999) that are typically available early in the project's life cycle regardless of the development process. The sizing measure in this study is defined as the sum of initial functional requirements plus initial external interfaces. This is pragmatic as these sizing artifacts are the only ones available in the DoD for budget and proposal evaluation.

This study will address the following research questions:

1. Do initial, as opposed to final, software requirements, defined as functional plus external interface requirements, relate to final agile development effort?
2. Do initial software requirements along with super domain, relate to final agile development effort? Super domain is defined as group of application domains with similar software characteristics. For example, vehicle payload and vehicle control are part of the same super domain called real-time embedded.
3. Do initial software requirements along with initial peak staff and super domain relate to final agile development effort?
4. Do initial software requirements relate to the final agile development schedule?
5. Do initial software requirements along with super domain relate to the final agile development schedule?
6. Do initial software requirements along with initial peak staff and super domain relate to final agile software development schedule?

Background

Motivation to Adopt Agile in the DoD

The recent adoption of agile in the DoD has been triggered by the need to move from a capabilities-based to a threat-based acquisition model to counter the rapid growing adversary capabilities. The DoD's traditional development process, based on upfront detailed system requirements for the entire completed system, is inadequate to meet these challenges. Senior officials believe that greater adoption of agile would result in significant improved acquisition performance and the ability to quickly respond to adversary technological advancements and update DoD software systems accordingly.

In 2009, the U.S. Congress enacted the National Defense Authorization Act for Fiscal Year 2010 (2010 NDAA) requiring the DoD to implement a new acquisition process for IT systems (United States House of Representatives , 2018). This new process included principles of agile development such as early and continual involvement of the user, multiple rapidly executed increments or releases, early successive prototyping to support an evolutionary approach, and a modular open-systems approach. Since the enactment of the 2010 NDAA, an increasing number of non-IT DoD acquisition programs have turned to agile software development as a method for delivering new and enhanced capabilities to the warfighters on a rapid and repeatable basis, avoiding delays and cost overruns associated with traditional methods such as waterfall (United States House of Representatives, 2018).

Software Development Processes in the DoD

Waterfall is the traditional software development process in the DoD (United States Department of Defense Science Board, 2018). Waterfall development [66] begins with writing down the full software requirements at the lowest level. Those software requirements are finalized and set by the government before contract award and documented in the software requirements specifications (SRS) and interface requirements specification (IRS). After contract award, the developer will use the government's full software requirements to write the program code as well as the test cases. When the software passes all test cases, it is considered finished and ready for delivery to the government.

Agile Software Development in the DoD

Agile development in the DoD is defined as a life cycle model that employs iterative and incremental development with active user engagement (DoD, 2020a). The main goal is to allow for continuous development throughout the software's life cycle (United States Department of Defense Science Board, 2018). It involves continuous planning, continuous testing, continuous integration, continuous feedback, and continuous evolution of the product. Software is developed in short iterations, called time boxes, which typically last one to four weeks. Each iteration is like a miniature software project of its own and includes all the traditional software activities (planning, requirements analysis, design, coding, testing, and documentation) to release the mini increment of new functionality. At the end of each iteration, the team reevaluates project priorities.

Conversely, **hybrid agile** combines principles of waterfall (United States Department of Defense Science Board, 2018) and agile development. That is, Waterfall for decomposing the software requirements for the entire system upfront; (2) Agile after contract award for segmenting, testing, and delivering the software in short iterations. This hybrid model is suitable for legacy systems (e.g., KC-46A Tanker) that are transitioning to agile, or in fixed-price contracts (e.g., technology demonstration and sustainment) where requirements are set or fully defined before award.

Scrum is the most widely used in the DoD as majority of agile projects are managed by small teams. However, DevSecOps (United States Department of Defense Chief Information Officer, 2019) is gradually becoming the preferred framework as new defense policies (DoD, 2020a) are aiming at applying security throughout the software life cycle in a cloud-based environment.

Agile Requirements Decomposition in the DoD

Most agile software projects in the DoD start by establishing high-level program goals and high-level software requirements (functional, non-functional, interfaces) and considered finished when the program goals have been met. Those high-level software requirements are written by the government (DoN, 2010) and documented in the SRS and IRS. After contract award, the developer will enter those high-level requirements in the product backlog, rewrite and decompose them into user stories, and continuously refine them (add, delete, modify) as small segments of software are developed and presented to the government for feedback.

Research Method

Instrumentation

The data collection questionnaire in the study was obtained from an existing one: *Software Resource Data Report (SRDR)* form (DoD, 2020b; Lipkin, 2011). SRDR is the primary source for actual software industrial data for the DoD and can be obtained by DoD analysts via the Cost Assessment Data Enterprise (CADE) repository (https://cade.osd.mil) owned by the Office of the Secretary of Defense for Cost Assessment and Program Evaluation (OSD CAPE). The SRDR is a regulatory contract reporting requirement (DoD, 2020b) for defense software developers.

The SRDR is used to obtain both the estimated and actual characteristics of new or upgrade software projects. The developer submits an Initial SRDR shortly after contract award and a Final SRDR after contract completion. These constitute a contract data deliverable for contractors that formalizes the reporting of software metrics and resource data. The SRDR questionnaire and data item description and form can be downloaded via the links below.

https://cade.osd.mil/content/cade/files/csdr/dids/current/dd3026-1_2019.XLSX

https://cade.osd.mil/Content/cade/files/csdr/guidance/DI-MGMT-82035A_SRDR%20Report.pdf

In the SRDR questionnaire, developers are required to indicate whether their development process is agile or hybrid agile (e.g., waterfall for Architecture and Requirements, followed by agile for design, code, and unit test).

In the SRDR questionnaire, developers are also required to report the total software requirements by category as shown in Table 2. The Initial SRDR includes the initial software requirements that were baselined at early phase. The Final SRDR includes the final software requirements at contract end. The final includes baselined plus changed requirements (added, modified, deleted) as a result of continuous refinements after award.

Table 14. Software Requirements Reported

Requirement Type	Initial SRDR	Final SRDR
Total Requirements	X	X
Functional Requirements		
Baselined	X	X
Added		X
Modified		X
Deleted		X
External Interface Requirements		
Baselined	X	X
Added		X
Modified		X
Deleted		X
Privacy Requirements	X	X
Security Requirements	X	X
Safety Requirements	X	X

Population and Sample

The sample includes 36 agile projects delivered for the DoD from 2008 to 2019. This study focused on completed agile software projects reported as computer software configuration items. The paragraphs below describe how these projects were characterized in terms of software requirements, operating environment, and super domain.

Operating Environment: The dataset was grouped into operating environments (Clark & Madachy, 2015; DoD, 2020b) as shown In Table 3. Operating environment is the host platform in which the developed software system operates. The dataset did not contain projects developed for space systems.

Table 15. Operating Environment

Type	Definition
Surface Fixed	Software is at a fixed site.
Air Vehicle	Software embedded as part of an aircraft or drone.
Sea System	software is embedded as part of a surface or underwater boat/ship or boat.
Ordnance System	Software embedded as part of a rocket or ordnance.
Missile System	Software embedded as part of a missile system

Super Domain: The dataset was grouped into super domains as shown in Table 4. The raw dataset was initially reported by application domains (Clark & Madachy, 2015; DoD, 2020b; Madachy et al., 2011; Rosa et al., 2014a, 2014b). The dataset was then stratified into four general complexity zones called super domains. This stratification was adopted from our recent work (Rosa et al., 2017). The application domains to super domain mapping are shown in Table 4.

Table 16. Super Domain Taxonomy

Super Domain	Definition	Application Domains included:
Support (SUPP)	Support software assists with operator training and software testing.	Software Tools Training
Automated Information Systems (AIS)	Provides information processing services to humans or other applications. Allows authority to exercise control and access to typical business processes.	Enterprise Resource Planning (ERP) Custom AIS Software Mission Planning
Engineering (ENG)	Take outputs of real-time software and further process them to provide human consumable information or automated control of devices.	Test Equipment Scientific Simulation Process Control System Software
Real-Time (RTE)	Most constrained type of software. These are specific solutions limited by system characteristics such as memory size, performance, or battery life. These take the most time and effort due to very high reliability or hardware constraints.	Communications Real Time Embedded Command & Control Vehicle Control Vehicle Payload Signal Processing Microcode/ Firmware

Variables in the Study

The variables examined are described in Table 5.

Table 17. Variable Names and Definitions

Variable	Type	Definition
Effort (E)	Dependent	Actual labor hours associated to all software activities: requirements analysis, architectural design, coding, Software Integration, Software Qualification Testing, Software Support Processes
Schedule (TDEV)	Dependent	▪ Actual development time (in months) to complete all software activities from Software Requirements Analysis through Software Qualification Test
Initial Software Requirements (REQ)	Independent	▪ Initial functional requirements + initial external interface requirements reported in the Initial SRDR Developer Report
Peak Staff (Staff)	Independent	Initial peak staff measured in terms of full-time equivalents reported in the ***Initial SRDR Developer Report.***
Super Domain	Categorical (Dummy)	Treatment of 4 (r) super domains required the addition of 3 (r-1) dummy variables : D1 = 1 if AIS, 0 if SUPP or otherwise D2 = 1 if ENG, 0 if otherwise D3 = 1 if RTE, 0 if otherwise

Model Selection

The model equation forms were chosen based on examining normal probability plots generated using the Cost Analysis Statistical Package (TECOLOTE Inc., 2020). The selection steps for the effort and schedule model forms are summarized below.

Figure 1 shows the residual plot for the linear regression of effort versus initial software requirements. The dataset does not appear to be normally distributed as residuals do not fall on the normality line. Consequently, lognormal regression was chosen for the three effort models.

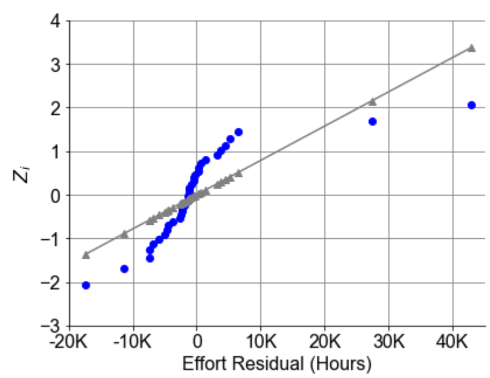


Figure 23. Effort Normal Probability Plot (Linear)

Figure 2 shows the residual plot for the linear regression of schedule versus initial software requirements. The dataset does not appear to be normally distributed as residuals do not fall on the normality line. Consequently, lognormal regression was chosen for the three schedule models.

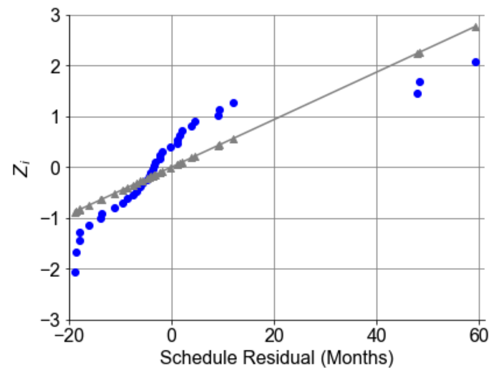


Figure 24. Schedule Normal Probability Plot (Linear)

Model Validation

The regression models were validated using the measures listed in Table 6.

Table 18. Model Validity Measures

Measure	Description
R^2	Coefficient of determination is the percentage of variation in the response explained by the model. (Matson et al., 1994)
R^2 (adj)	Adjusted R2 is the percentage of the variation in the response explained by the model, adjusted for the number of predictors in the model relative to the number of observations.
R^2 (pred)	Predicted R2 is a cross validation method that involves removing each observation from the dataset, estimating the regression equation, determining how well the model predicts the removed observation, and repeats this for all data points.
P-value	Statistical significance through the coefficient alpha (α = 0.05).
VIF	Variance Inflation Factor indicates if multicollinearity is present in a multi-regression analysis; VIF lower than 10, indicates no multicollinearity.
SEE	Standard Error of the Estimate is the difference between observed and the estimated effort. SEE is to linear models as standard deviation is to sample means.
F-test	F test is the square of the equivalent t test; the bigger it is, the smaller the probability that difference could occur by chance.
MMRE	Mean Magnitude of Relative Error is an indicator of model's accuracy: Low MMRE= high accuracy. (Mukhopadhyay & Kekre, 1992)

Dataset Demographics

The sample was identified as 36 agile projects completed from 2008 to 2019 (Figure 3), involving six operating environments (Figure 4), and four super domains (Figure 5). The majority of the projects were completed in the last six years and most of the software projects were hosted at a surface fixed site (e.g., mission operations center, data center).

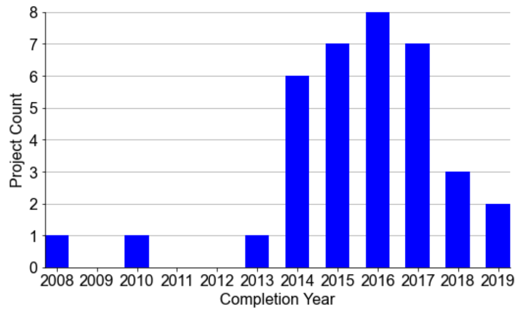


Figure 25. Agile Project Completion Year

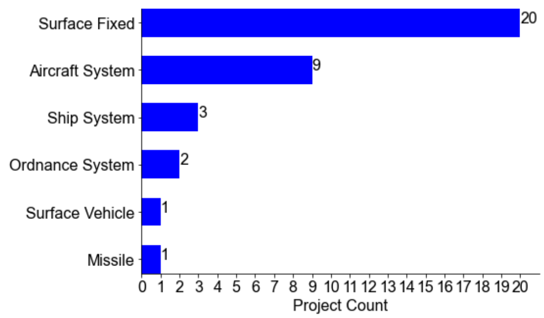


Figure 26. Operating Environment

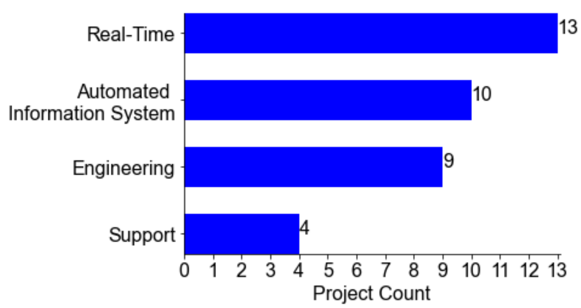


Figure 27. Super Domain

Figure 6 displays the agile process for the 36 projects. Projects were characterized as agile or hybrid agile. The projects identified as **hybrid agile** used waterfall process for requirements analysis, and agile process for design, code, unit test, and integration. This information was obtained from developer's data item descriptions provided as a supplement to their SRDR submission. We also contacted the developers to validate and verify their responses.

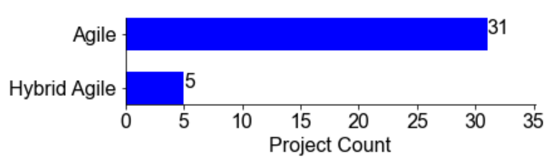


Figure 28. Agile Process

Figure 7 displays the agile framework for the dataset. The majority of the projects used Scrum. This information was obtained from developer's SRDR questionnaire, data item description, and proprietary documents describing their processes. We also contacted the developers to validate and verify their responses.

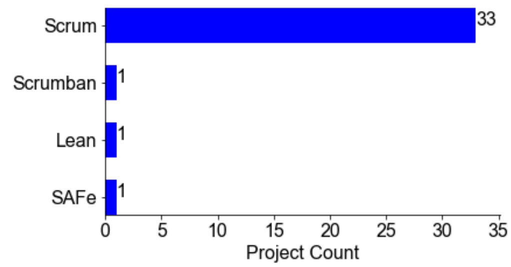


Figure 29. Agile Framework

Descriptive Statistics

Figure 8 is a histogram of the actual software development effort for the agile projects. The average software development effort for the sample was 99,959 hours and standard deviation of 134,641. The distribution appears to be right skewed, confirming the non-normal distribution of effort data as previously shown in Figure 1.

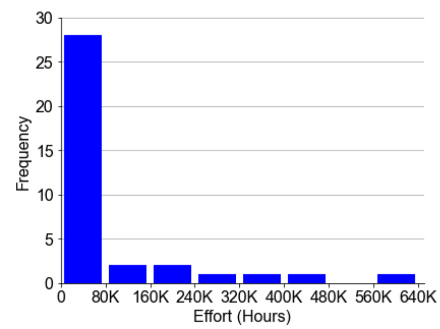


Figure 30. Effort Histogram

Figure 9 is a histogram of the actual development time for the agile projects. The average software development completion time for the sample was 26.95 months and standard deviation of 19.73. The distribution again confirms the non-normal distribution of the data as previous shown in Figure 2.

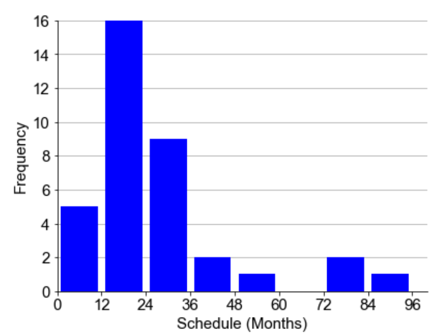


Figure 31. Schedule Histogram

Figure 10 provides a histogram of initial software requirements. The average total software requirements for the sample was 798. The distribution is lognormal.

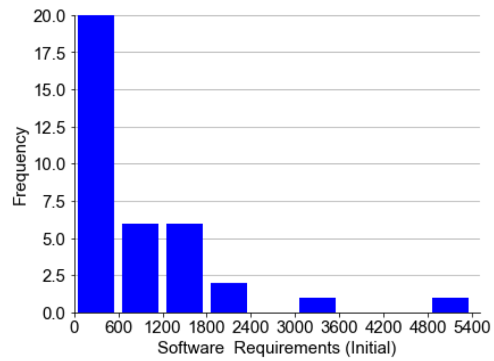


Figure. 32. Software Size Histogram

Figure 11 shows a histogram of peak staff for the agile project dataset showing Full-Time Equivalent (FTE) staff. The average peak staff for the sample was 31. The project with largest peak staff was developed using SAFe. This data has a lognormal distribution.

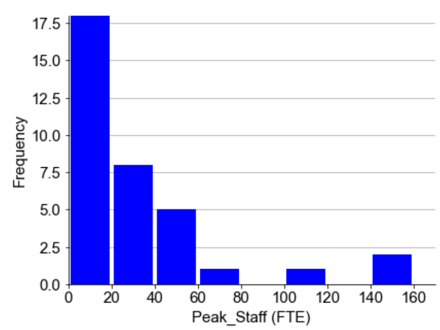


Figure 11. Peak Staff Histogram

Figure 12 provides a histogram of the requirements volatility (RVOL) for the agile project dataset. RVOL is the sum of requirements changes (added, modified, and deleted) divided by the total number of baselined requirements. Baselined requirements are those initial software requirements used in this study. The average RVOL for the sample was 19%. More than half experienced RVOL ≥ 12%. This confirms the notion that software requirements for agile projects in DoD are not fixed and evolve over time.

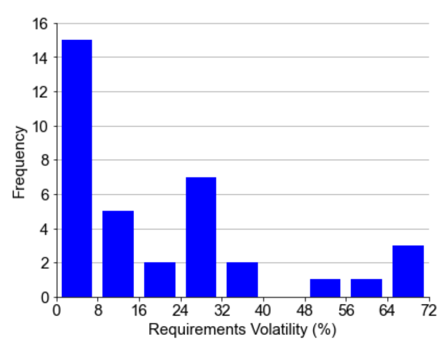


Figure 33. Requirements Volatility Histogram

Results

This section describes the resulting effort and schedule models associated with Research Questions (RQ) 1 through 6. Loglinear regression was performed for all models using the Cost Analysis Statistical Package (TECOLOTE Inc., 2020). Log-normal ordinary least squares (OLS) regression was used to create the multiplicative models in this study. The data is transformed into log-space and the coefficients are derived using OLS regression. The derived coefficients for each predictor variable are treated as exponents and the regression intercept is transformed back into unit-space with an anti-log. Alpha (α = 0.05) is used in evaluating the p-values for each model. These models are applicable for DoD agile software project size ranging between 10 and 5000 initial software requirements, and a peak staff between 1 and 150 full-time equivalents. The sample size, however, may impact models' ability to detect statistical effects with any greater power.

Effort Model 1

RQ 1: Do initial, as opposed to final, software require-ments, defined as functional plus external interface requirements, relate to final agile development effort?

Equation (1) predicts effort for agile software development projects as a function of total initial software requirements.

$$\mathrm{E} = 1006 x REQ^{0.65} \quad (1)$$

Where:

E	=	Final software development hours
REQ	=	FUNC + EIF

And:

FUNC	=	Initial Functional Requirements
EIF	=	Initial External Interface Requirements

Figure 13 shows the residual plot for equation (1). The residuals approximate a straight line. This suggests that loglinear regression is valid for modeling.

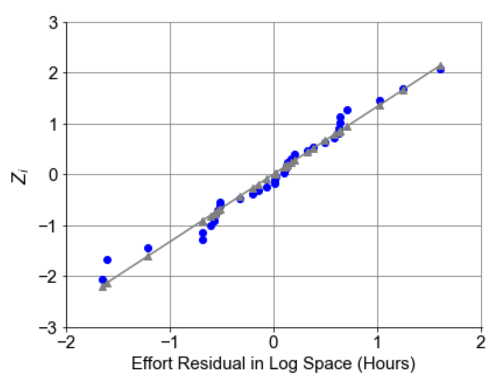


Figure 34. Normal Probability Plot for Equation (1)

Table 7 reports the coefficient statistics, goodness-of-fit test, and analysis of variance for

Equation (1). The high t-statistics and low p-value suggest that initial software requirements are strongly correlated to effort. The result also adds insight to the fact that initial functional requirements plus initial external interface requirements are effective in estimating effort for agile projects. The small difference between adjusted and predicted R^2 suggest that the model predicts new observations as well as it fits the existing data. However, the low adjusted R^2 of 63% suggests adding additional variables for a better model fit.

Table 19. Regression Analysis Results for Equation (1)

Coefficient Statistics Summary

Term	Coef	T-Statistic	P-value	VIF
Intercept	6.9138	13.7	0.0000	***
REQ	0.6500	7.8	0.0000	***

Goodness-of-Fit Test

SE	R^2	R^2 (adj)	R^2 (pred)	MMRE
0.7274	64.11%	63.06%	60.29%	68.16%

Analysis of Variance

Source	DF	Sum of Sq.	Mean Sq.	F-stat
Regression	1	32.14	32.14	60.74
Residual	34	17.99	0.52	
Total	35	50.13		

Effort Model 2

RQ 2: Do initial software requirements along with super domain, relate to final agile development effort?

Equation (2) predicts effort for agile software development projects using total initial software requirements as predictor and super domain as categorical variables (D1, D2, D3).

$$E = 200.7 x REQ^{0.7182}(3.0^{D1})(3.6^{D2})(5.1^{D3}) \qquad (2)$$

Where:

E = Final development labor hours

REQ = FUNC + EIF

And:

FUNC = Initial Functional Requirements

EIF = Initial External Interface Requirements

D1 = 1 if AIS, 0 if otherwise

D2 = 1 if ENG, 0 if otherwise

D3 = 1 if RTE, 0 if otherwise

If D1, D2 and D3 are zero, Equation (2) predicts effort for the SUPP super domain

Figure 14 shows the normal residual plot for Equation (2). Loglinear regression is valid for this model as its residuals approximate a normal distribution.

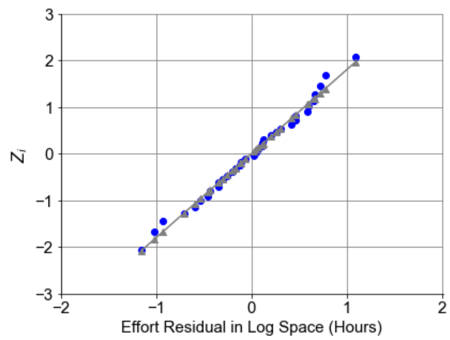


Figure 35. Normal Probability Plot for Equation (2)

Table 8 shows the analysis results for Equation (2). The t-statistics and p-values suggest that super domain is strongly correlated to effort, and low VIF values indicates no sign of multicollinearity. The small difference between adjusted and predicted R^2 also suggest the model predicts new observations as well as it fits the existing data. This model shows higher adjusted R^2 and lower MMRE than Equation (1); signifying that adding super domain categorical variables to Equation (1) improves accuracy and fit.

Table 20. Regression Analysis Results for Equation (2)

Coefficient Statistics Summary

Term	Coef	T-Statistic	P-value	VIF
Intercept	5.3019	9.7	0.0000	
REQ	0.7182	10.2	0.0000	1.2
D1	1.0929	3.2	0.0028	2.5
D2	1.2728	3.5	0.0013	2.7
D3	1.6332	4.9	0.0000	2.8

Goodness-of-Fit Test

SE	R^2	R^2 (adj)	R^2 (pred)	MMRE
0.5676	80.08%	77.51%	73.60%	47.52%

Analysis of Variance

Source	DF	Sum of Sq.	Mean Sq.	F-stat
Regression	4	40.14	10.0366	31.14
Residual	31	9.98	0.3222	
Total	35	50.13		

Effort Model 3

RQ3: Do initial software requirements along with initial peak staff and super domain relate to final agile development effort?

Equation (3) predicts software development effort for agile projects using peak staff and total initial software requirements as predictors, while super domain as categorical variables (D1, D2, D3).

$$E = 173.2xREQ^{0.539}Staff^{0.463}(2.3^{D1})(3.7^{D2})(3.9^{D3}) \quad (3)$$

Where:

E = Final development labor hours

REQ = FUNC + EIF

Staff = Initial peak staff in full-time equivalent

And:

FUNC = Initial Functional Requirements
EIF = Initial External Interface Requirements
D1 = 1 if AIS, 0 if otherwise
D2 = 1 if ENG, 0 if otherwise
D3 = 1 if RTE, 0 if otherwise

If D1, D2 and D3 are zero, Equation (3) predicts effort for the SUPP super domain

Figure 15 shows the transformed normal residual plot for Equation (3). Loglinear regression is valid for this model as its residuals approximate a straight line.

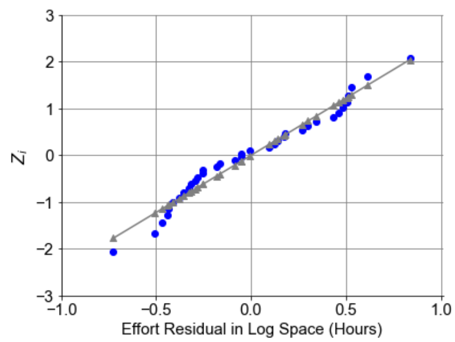


Figure 36. Normal Probability Plot for Equation (3)

Table 9 shows the statistical analysis for Equation (3). The t-statistics suggests that all three variables are strongly correlated to effort; with no signs of multicollinearity. The small difference between adjusted and predicted R^2 suggest the model predicts new observations as well as it fits the existing data. This model shows higher adjusted R^2 and lower MMRE than Equations (1) and (2) signifying that this model achieves the highest accuracy and best fit when all three variables are added to the regression.

Table 21. Regression Analysis Results for Equation (3)

Coefficient Statistics Summary

Term	Coef	T-Statistic	P-value	VIF
Intercept	5.1543	12.7	0.0000	
REQ	0.5390	8.6	0.0000	1.7
Staff	0.4631	5.2	0.0000	1.8
D1	0.8362	3.3	0.0025	2.6
D2	1.3025	4.9	0.0000	2.7
D3	1.3696	5.5	0.0000	2.9

Goodness-of-Fit Test

SE	R^2	R^2 (adj)	R^2 (pred)	MMRE
0.4198	89.46%	87.70%	84.74%	33.85%

Analysis of Variance

Source	DF	Sum of Sq.	Mean Sq.	F-stat
Regression	5	44.8489	8.9698	50.90
Residual	30	5.2861	0.1762	
Total	35	50.1350		

Schedule Model 4

RQ4: Do initial software requirements relate to the final agile development schedule?

Equation (4) predicts software development time (in months) for agile projects as a function of total initial software requirements.

$$TDEV = 6.84xREQ^{0.202} \quad (4)$$

Where:

TDEV = Final development time (in months)

REQ = FUNC + EIF

And:

FUNC = Initial Functional Requirements

EIF = Initial External Interface Requirements

Figure 16 shows the transformed normal residual plot for Equation (4). Loglinear regression is valid for this model as its residuals follow approximate a normal distribution.

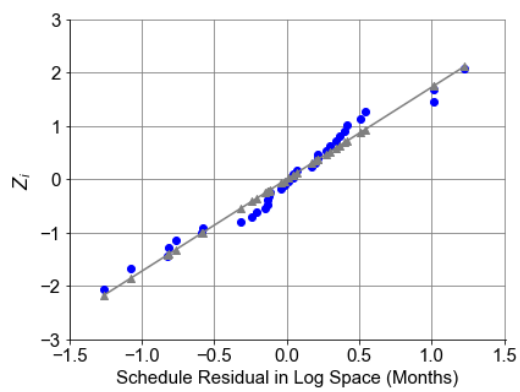


Figure 37. Normal Probability Plot for Equation (4)

Table 10 reports the coefficient statistics, goodness-of-fit test, and analysis of variance for Equation (4). The high t-statistics and low p-value suggests that initial software requirements are strongly correlated to development time (TDEV). However, the low R^2 is an indication that a schedule model only based on initial software requirements does not fit the data well. Adding additional predictors to the model may increase the R^2.

Table 22. Regression Analysis Results for Equation (4)

Coefficient Statistics Summary

Term	Coef	T-Statistic	P-value	VIF
Intercept	1.8998	4.89	0.0000	***
REQ	0.2029	3.16	0.0033	***

Goodness-of-Fit Test

SE	R^2	R^2 (adj)	R^2 (pred)	MMRE
0.5598	22.71%	20.44%	14.53%	46.66%

Analysis of Variance

Source	DF	Sum of Sq.	Mean Sq.	F-stat
Regression	1	3.1306	3.1306	9.99
Residual	34	10.6543	0.3134	
Total	35	13.7849		

Schedule Model 5

RQ 5: Do initial software requirements along with super domain relate to final agile development schedule?

Equation (5) predicts software development time (TDEV) for agile projects using total initial software requirements as predictor and super domain as categorical variable (D1, D2, D3).

$$TDEV = 1.64xREQ^{0.272}(2.1^{D1})(2.9^{D2})(4.0^{D3}) \quad (5)$$

Where:

TDEV = Final development time (in months)

REQ = FUNC + EIF

And:

FUNC = Initial Functional Requirements

EIF = Initial External Interface Requirements

D1 = 1 if AIS, 0 if otherwise

D2 = 1 if ENG, 0 if otherwise

D3 = 1 if RTE, 0 if otherwise

If D1, D2 and D3 are zero, Equation (5) predicts effort for the SUPP super domain

Figure 17 shows the transformed normal residual plot for Equation (5). Loglinear regression is valid for this model as its residuals approximate a normal distribution.

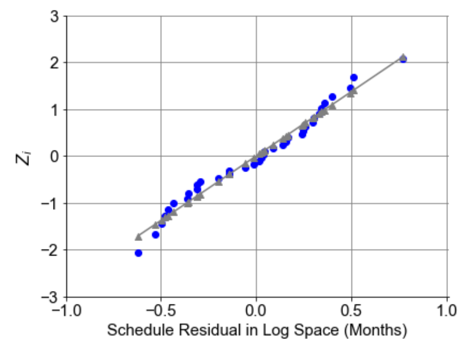


Figure 38. Normal Probability Plot for Equation (5)

Table 11 shows the analysis results for Equation (5). The t-statistics and p-values suggest that super domain is strongly correlated to effort, and low VIF values indicate no multicollinearity. The small difference between adjusted and predicted R^2 indicates the model may predict new observations as well as it fits the existing data. This model shows higher adjusted R^2 (20% → 65%) and lower MMRE (46%→ 30%) compared to Equation (4); signifying that adding super domain categorical variables improves its accuracy and fit.

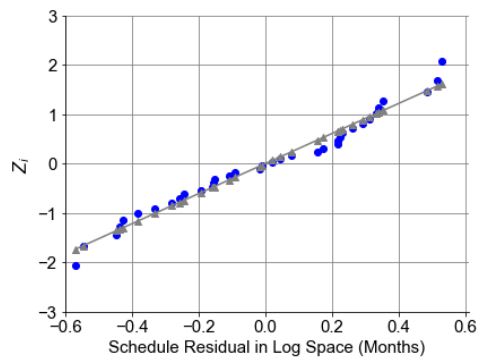


Figure 39. Normal Probability Plot for Equation (6)

Table 23. Regression Analysis Results for Equation (5)

Coefficient Statistics Summary

Term	Coef	T-Statistic	P-value	VIF
Intercept	0.4980	1.40	0.1724	
REQ	0.2722	5.97	0.0000	1.2
D1	0.7639	3.49	0.0015	2.5
D2	1.0972	4.69	0.0001	2.7
D3	1.4072	6.56	0.0000	2.8

Goodness-of-Fit Test

SE	R^2	R^2 (adj)	R^2 (pred)	MMRE
0.3691	69.37%	65.42%	59.14%	30.07%

Analysis of Variance

Source	DF	Sum of Sq.	Mean Sq.	F-stat
Regression	4	9.5625	2.3906	17.5514
Residual	31	4.2224	0.1362	
Total	35	13.7849		

Schedule Model 6

RQ6: Do initial software requirements along with initial peak staff and super domain relate to final agile software development schedule?

Equation (6) predicts software development time (TDEV) for agile projects using peak staff and total initial software requirements as predictors, while super domain as categorical variables (D1, D2, D3).

$TDEV = 1.74xREQ^{0.345}Staff^{-0.189}(2.3^{D1})(3.0^{D2})(4.5^{D3})$ (6)

Where:

TDEV = Final development time (in months)

REQ = FUNC + EIF

Staff = Initial peak staff in full-time equivalent

And:

FUNC = Initial Functional Requirements

EIF = Initial External Interface Requirements

D1 = 1 if AIS, 0 if otherwise

D2 = 1 if ENG, 0 if otherwise

D3 = 1 if RTE, 0 if otherwise

If D1, D2 and D3 are zero, Equation (5) predicts effort for the SUPP super domain.

Figure 18 shows the transformed normal residual plot for Equation (6). Loglinear regression is valid for this model as its residuals approximate a normal distribution.

Table 12 shows the statistical analysis for Equation (6). The t-statistics shows all three variables are strongly correlated to TDEV; with no signs of multicollinearity. The small difference between adjusted and predicted R^2 suggests the model predicts new observations as well as it fits the existing data. This model shows higher adjusted R^2 and lower MMRE than Equations (4) and (5); suggesting that this model achieves highest accuracy and best fit.

Table 24. Regression Analysis Results for Equation (6)

Coefficient Statistics Summary

Term	Coef	T-Statistic	P-value	VIF
Intercept	0.5585	1.7	0.0986	
REQ	0.3456	6.9	0.0000	1.7
Staff	-0.1896	-2.6	0.0135	1.8
D1	0.8690	4.2	0.0002	2.6
D2	1.0850	5.1	0.0000	2.7
D3	1.5151	7.5	0.0000	2.9

Goodness-of-Fit Test

SE	R^2	R^2 (adj)	R^2 (pred)	MMRE
0.3383	75.09%	70.94%	63.24%	27.50%

Analysis of Variance

Source	DF	Sum of Sq.	Mean Sq.	F-stat
Regression	5	10.3508	2.0702	18.08
Residual	30	3.4340	0.1145	
Total	35	13.7849		

Discussion of Results

The resulting statistics add insight to the notion that initial, as opposed to final, software requirements, when defined as functional and external interface requirements in the SRS and IRS, is good at predicting effort and time for DoD agile projects.

The multi-variable models ((3) and (6)) based on requirements, peak staff and super domain as inputs, perform better than single-variable models ((1) and (4)) using requirements alone to predict effort or schedule.

Model Usefulness and Limitations

The models in this study are useful for effort and schedule estimates at proposal evaluation or before. Since these models were derived using OLS in log-space (using the natural log), the output represents an estimate at the 50% conficence level in log-space. To understand the uncertainty in the different models, the model results should be displayed as a 95% conficence interval rather than a single value. The confidence interval is derived in log-space using two times the model's standard error (SEE). For example, the 95% confidence interval for the output from Equation (3) and the RTE super domain is expressed as follows:

$$\text{E (Low)} = 173.2xREQ^{0.539}Staff^{0.463}x(3.9) - (2 \times .4198)$$

$$\text{E (High)} = 173.2xREQ^{0.539}Staff^{0.463}\ x(3.9) + (2 \times .4198)$$

The result is transformed into unit-space by taking the natural anti-log. The confidence intervals derived from these models can help program managers independently assess whether their software development contract is in breach status. For example, if the contract's latest revised schedule or cost estimate exceeds the prediction interval's upper bound, the acquisition decision authority may declare a contract breach or restructure the program.

Threats to Validity

Possible threats to validity are summarized next:

Internal Validity Threats:

- The dataset timeframe (2008–2019) raises potential issues as the earlier projects (2008, 2010, 2013) were developed using agile processes tailored to fit the developer's need. It is likely that agile processes evolved during the 11-year timeframe.
- This study is on frameworks commonly considered "agile" and a focus on only one of the frameworks may produce different results.
- The developers "self-reported" the data in the SRDR questionnaire. We verified 80% (29 out of 36) of the project data by contacting the developers and following up for additional information. Whether the unverified projects (20%) are accurate remains unanswered.

External Validity Threats:

- The data from this study came from DoD contracts that exceeded $50 million in value for the total contract. The performance of these large companies may not be generalizable to other organizations.
- The initial software requirement counts came from SRS and IRS, a common artifact in DoD acquisition. Non-DoD organizations may not use an SRS to state their requirements at early phase.
- These models proved to be effective in estimating total development hours and duration for agile projects reported at the release level in the DoD. However, we cannot generalize

beyond this group for several reasons. First, majority of the projects were developed using Scrum and none in XP. Second, the initial software requirements were developed at a high-level and only included functions and interfaces. Third, models may not be suitable for projects using DevOps as the reported effort in the dataset does not capture sustainment activities.

Construct Validity Threats:

- A sample size of 36 agile projects poses a threat to statistical conclusion as it does not allow for detecting effects with greater power. A larger sample size is needed for confirmatory hypothesis testing.
- Projects in this study reported their initial software requirements using the traditional SRS and IRS templates. After contract award, those initial requirements were rewritten into stories and continuously refined using agile processes. Whether the projects should be classified as agile or hybrid agile remains debatable.

Example Applications

The effort and schedule models are used here for estimating two examples: a Radar Display Manager and a Testing Tool. The examples show how to take information on requirements, peak staffing, and the application super domain and use it in the models. Of particular interest is how super domains are represented in each model as there are four super domains but only three super domain variables in the effort and schedule models.

Radar Display Manager

In the first example, there is a need to estimate the software development effort and schedule for Radar Display Manager software. There are 207 Initial Requirements and 23 External Interface Requirements for a total of 230 requirements. The estimated Peak Staffing is 16 people.

The Radar Display Manager is in the Real Time Embedded (RTE) super domain. Effort Model (3) takes requirements, peak staffing, and super domain as inputs. Super domains are represented by the D1, D2, and D3 variables for the AIS, ENG, and RTE super domains respectively. The super domain variables have a value of either zero (0) or one (1). In this example, the Radar Display Manager is in the RTE super domain, variable D3 is set to 1 and the other two variables are set to 0.

Therefore, the effort estimation model for the Radar Display Manager is:

$$E = 173.2\ x\ 230^{0.539}\ x\ 16^{0.463}\ x\ (2.3^{0})\ x\ (3.7^{0})\ x\ (3.9^{1})$$

$$E = 173.2\ x\ 18.75\ x\ 3.6\ x\ 1\ x\ 1\ x\ 3.9$$

$$E = 45{,}595\ hours$$

For the Radar Display Manager software, the time to develop (TDEV) schedule estimation model is:

$$TDEV = 1.74\ x\ 230^{0.345}\ x\ 16^{-0.189}\ x\ (2.3^{0})\ x\ (3.0^{0})\ x\ (4.5^{1})$$

$$TDEV = 1.74\ x\ 6.53\ x\ 0.59\ x\ 1\ x\ 1\ x\ 4.5$$

$$TDEV = 30\ months$$

The estimate for the Radar Display Manager is 45,595 hours and 30 months. That is an average of 1,520 hours per month.

Testing Tool

In the next example, the effort and schedule need to be estimated for the software development of a Testing Tool. There are 31 Initial Requirements and five External Interface Requirements for a total of 36 requirements. The estimated Peak Staffing is four people.

The Testing Tool is in the Support (SUPP) super domain. There are variables in the effort and schedule models for the AIS (D1), ENG (D2), and RTE (D3) super domains but no variable for the SUPP super domain. This is because the base models for effort and schedule are for the SUPP super domain when the D1, D2, and D3 variables are given the value of zero (0).

The Testing Tool effort estimation model is represented as:

$$E = 173.2\ x\ 36^{0.539}\ x\ 4^{0.463}\ x\ (2.3^{0})\ x\ (3.7^{0})\ x\ (3.9^{0})$$

$$E = 173.2\ x\ 6.9\ x\ 1.9\ x\ 1\ x\ 1\ x\ 1$$

$$E = 2{,}271$$

The TDEV schedule estimation model for the Testing Tool is:

$$TDEV = 1.74\ x\ 35^{0.345}\ x\ 4^{-0.189}\ x\ (2.3^{0})\ x\ (3.0^{0})\ x\ (4.5^{0})$$

$$TDEV = 1.74\ x\ 3.44\ x\ 0.77\ x\ 1\ x\ 1\ x\ 1$$

$$TDEV = 4.6\ months$$

The estimate for the Testing Tool software is 2,271 hours, 4.6 months, and 494 hours per month.

Conclusion

The results add insight to the notion that initial, ***as opposed to final***, functional plus external interface requirements, when treated as sizing input along with a super domain categorical variable, are effective in predicting software development effort and schedule of DoD agile projects early in the life cycle; at the time when mainstream agile sizing metrics are not available for estimation in DoD and budget/schedule baselines are being established. These models can be used for building independent government cost estimates to crosscheck request for proposals.

The results also suggest that DoD contractors should consider adding peak staff along with initial software requirements (as defined in this study) and super domain as inputs when building effort and schedule models for their agile project cost proposal. These three inputs are generally obtained from: contract proposals (i.e., peak staff), request for proposals (i.e., super domain), and government provided requirements (i.e., initial software requirements).

References

Abraham, S. & Insfran, E. (2008, September 3–5). A metamodeling approach to estimate software size from requirements specifications. *Software Engineering and Advanced Applications, 2008. SEAA '08. 34th Euromicro Conference*, 465–475.

Alleman, G. B., Henderson, M. & Seggelke, R. (2003). Making Agile development work in a government contracting environment-measuring velocity with earned value. *Proceedings of the Agile Development Conference, 2003. ADC 2003*, 114–119.

Alshayeb, M. & Li, W. (2003, November). An empirical validation of object-oriented metrics in two different iterative software processes. *IEEE Transactions on Software Engineering*, *29*(11), 1043–1049.

Bilgaiyan, S., Mishra, S. & Das, M. (2016). A review of software cost estimation in Agile software development using soft computing techniques. *2016 2nd International Conference on Computational Intelligence and Networks (CINE)*, 112–117.
Choetkiertikul, M., Dam, H. K., Tran, T., Ghose, A. & Grundy, J. (2018, June 1). Predicting delivery capability in iterative software development. *IEEE Transactions on Software Engineering, 44*(6), 551–573.
Choetkiertikul, M., Dam, H. K., Tran, T., Pham, T., Ghose, A. & Menzies, T. (2019, July 1). A deep learning model for estimating story points. *IEEE Transactions on Software Engineering*, *45*(7), 637–656.
Clark, B. & Madachy, R. (Eds.). (2015, April). *Software cost estimation metrics manual for defense systems*. Software Metrics Inc. http://www.sercuarc.org/wp-content/uploads/2014/05/Software-Cost-Estimation-Metrics-Manual-for-Defense-Systems.pdf
Department of Defense. (2016, March). *Agile and earned value management: A program managers desk guide*. http://www.acq.osd.mil/evm/docs/PARCA
Department of Defense. (2020b, October). *DoD instruction 5000.87, operation of the software acquisition pathway*. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500087p.PDF?ver=virAfQj4v_LgN1JxpB_dpA%3d%3d
Department of Defense. (2020b, November). *Software resource data report*. https://cade.osd.mil/content/cade/files/csdr/dids/current/dd3026-1_2019.XLSX
Department of Navy. (2010, September). *Software criteria and guidance for systems engineering technical reviews (SETR) supplement to guidebook for acquisition of Naval software intensive systems*. http://www.secnav.navy.mil/rda/OneSource/Documents/Program%20Assistance%20and%20Tools/Handbooks,%20Guides%20and%20Reports/Page%205/supplementtoguidebook.pdf
Ebrahimi, N. B. (1999, January–February). How to improve the calibration of cost models. *IEEE Transactions on Software Engineering, 25*(1), 136–140.
GAO. (2012, July). *Software development: Effective practices and federal challenges in applying Agile methods*. http://www.gao.gov/assets/600/593091.pdf
GAO. (2020, September). *Agile assessment guide: Best practices for Agile adoption and implementation* (GAO-20-590C). https://www.gao.gov/assets/710/709711.pdf
Garg, S., & Gupta, D. (2015). PCA based cost estimation model for agile software development projects. *2015 International Conference on Industrial Engineering and Operations Management (IEOM)*, 1–7.
Hastings, T. E. & Sajeev, A. S. M. (2001, April). A vector-based approach to software size measurement and effort estimation. *IEEE Transactions on Software Engineering, 27*(4), 337–350.
Heričko, M., & Živkovič, A. (2008, June). The size and effort estimates in iterative development. *Information and Software Technology, 50*(7–8), 772–781. http://www.sciencedirect.com/science/article/pii/S0950584907000870
Jones, C. (2013). Function points as a universal software metric. *SIGSOFT Software Engineering Notes 38, 4*(July 2013), 1–27.
Jorgensen, M. (2005, November). Evidence-based guidelines for assessment of software development cost uncertainty. *IEEE Transactions on Software Engineering, 31*(11), 942–954.
Jorgensen, M. & Grimstad, S. (2011, September–October). The impact of irrelevant and misleading information on software development effort estimates: A randomized controlled field experiment. *IEEE Transactions on Software Engineering, 37*(5), 695–707.
Kang, S., Choi, O. & Baik, J. (2010). Model-based dynamic cost estimation and tracking method for Agile software development. *2010 IEEE/ACIS 9th International Conference on Computer and Information Science*, 743–748.
Kaya, M., & Demirors, O. (2011). E-Cosmic: A business process model based functional size estimation approach. *Software Engineering and Advanced Applications (SEAA), 2011 37th EUROMICRO Conference on, Inf. Inst.*, 404–410.
Lipkin, I. (2011). *Test software development project productivity model* [Doctoral dissertation, University of Toledo].

Madachy, R., Boehm, B., Clark, B., Tan, T., & Rosa, W. (2011). U.S. DoD application domain empirical software cost analysis. *2011 International Symposium on Empirical Software Engineering and Measurement,* 392–395.

Malik, A. & Boehm, B. (2011, July). Quantifying requirements elaboration to improve early software cost estimation. *Information Sciences, 181(*13) 2747–2760. http://www.sciencedirect.com/science/article/pii/S002002550900526X

Malik, A. (2011). *Quantitative and qualitative analyses of requirements elaboration for early software size estimation* [Doctoral dissertation, University of Southern California].

Matson, J. E., Barrett, B. E., & Mellichamp, J. M. (1994, April). Software development cost estimation using function points. *IEEE Transactions on Software Engineering, 20*(4), 275–287.

Molokken-Ostvold, K. & Jorgensen, M. (2005, September). A comparison of software project overruns—flexible versus sequential development models. *IEEE Transactions on Software Engineering, 31*(9), 754–766.

Mukhopadhyay, T. & Kekre, S. (1992, October). Software effort models for early estimation of process control applications. *IEEE Transactions on Software Engineering, 18*(10), 915–924.

Nan, N. & Harter, D. E. (2009, September–October). Impact of budget and schedule pressure on software development cycle time and effort. *IEEE Transactions on Software Engineering, 35*(5), 624–637.

Nassif, A., Ho, D., & Capretz, L. (2013, January). Towards an early software estimation using log-linear regression and a multilayer perceptron model. *Journal of Systems and Software, 86*(1), 144–160. http://www.sciencedirect.com/science/article/pii/S0164121212002221

Nguyen-Cong, D. & Tran-Cao, D. (2013). A review of effort estimation studies in Agile, iterative and incremental software development. *The 2013 RIVF International Conference on Computing & Communication Technologies—Research, Innovation, and Vision for Future (RIVF)*, 27–30.

Ochodek, M., Nawrocki, J., & Kwarciak, K. (2011, March). Simplifying effort estimation based on use case points. *Information and Software Technology, 53(*3), 200–213. http://www.sciencedirect.com/science/article/pii/S095058491000176X

Owais, M. & Ramakishore, R. (2016). Effort, duration and cost estimation in agile software development. *2016 Ninth International Conference on Contemporary Computing (IC3), Noida*, 1–5.

Pendharkar, P. C., Subramanian, G. H., & Rodger, J. A. (2005, July). A probabilistic model for predicting software development effort. In *IEEE Transactions on Software Engineering, 31*(7), 615–624.

Pillai, K. & Sukumaran Nair, V. S. (1997, August). A model for software development effort and cost estimation. In *IEEE Transactions on Software Engineering, 23*(8), 485–497.

Popli, R. & Chauhan, N. (2013). A sprint-point based estimation technique in Scrum. *2013 International Conference on Information Systems and Computer Networks, Mathura*, 98–103.

Popli, R. & Chauhan, N. (2014). Agile estimation using people and project related factors. *2014 International Conference on Computing for Sustainable Global Development (INDIACom)*, 564–569.

Popli, R. & Chauahn, N. (2015). Managing uncertainty of story-points in Agile software. *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, 1357–1361.

Rosa, W. , Madachy, R., Boehm, B., & Clark, B. (2014a). Simple empirical software effort estimation model. *ESEM '14 Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement,* Article No. 43.

Rosa, W., Clark, B., Boehm, B., & Madachy, R. (2014b, October). Simple-empirical software cost estimation. ***29th International Forum on COCOMO and Systems/Software Cost Modeling [Online].*** http://csse.usc.edu/new/wp-content/uploads/2014/10/Simple-Empirical-Software-Cost-Estimation_v2.pdf

Rosa, W., Madachy, R., Clark, B., & Boehm, B. (2017). Early phase cost models for Agile software processes in the U.S. DoD. *2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM),* 30–37. doi:10.1109/ESEM.2017.10 http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8170082&isnumber=8169971

Saini, A., Ahuja, L., & Khatri, S. K. (2018). Effort estimation of Agile development using fuzzy logic. *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO),* 779–783.

Sharma, A., & Kushwaha, D. (2010, October 5). Early estimation of software complexity using requirement engineering document. *SIGSOFT Software Engineering Notes,* 35.

Sharma, A., & Kushwaha, D. (2012). Estimation of software development effort from requirements based complexity. *Procedia Technology, 4*, 716–722.

Shepperd, M. & Schofield, C. (1997, November). Estimating software project effort using analogies. *IEEE Transactions on Software Engineering, 23*(11), 736–743.

Tanveer, B., Guzmán, L., & Engel, U. M. (2016). Understanding and improving effort estimation in agile software development—An industrial case study. *2016 IEEE/ACM International Conference on Software and System Processes (ICSSP),* 41–50.

TECOLOTE Inc. (2020, November). Automated cost estimating integrated tool: CO$TAT. https://www.aceit.com/aceit-suite-home/product-info/costat

United States Department of Defense Chief Information Officer. (2019, September). *DoD enterprise DevSecOps reference design*. https://dodcio.defense.gov/Portals/0/Documents/DoD%20Enterprise%20DevSecOps%20Reference%20Design%20v1.0_Public%20Release.pdf?ver=2019-09-26-115824-583

United States Department of Defense Science Board. (2018, February). *Design and acquisition of software for defense systems.* https://apps.dtic.mil/dtic/tr/fulltext/u2/1048883.pdf

United States House of Representatives. (2018, June). *Department of Defense appropriations bill, 2019*. https://slidelegend.com/queue/department-of-defense-appropriations-bill-2019-pdf-_5b6a213c097c47ac1b8b45bc.html#

Usman, M., & Britto, R. (2016). Effort estimation in co-located and globally distributed Agile software development: A comparative study. *2016 Joint Conference of the International Workshop on Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA)*, 219–224.

Usman, M., Mendes, E., Weidt, R., & Britto, R. (2014). Effort estimation in Agile software development: A systematic literature review. In *Proceedings of the 10th International Conference on Predictive Models in Software Engineering* (PROMISE '14), 82–91. ACM, New York, NY, USA.

Usman, M., Mendes, E., & Börstler, J. (2015). Effort estimation in Agile software development: A survey on the state of the practice. In *Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering* (EASE '15). ACM, New York, NY, USA.

Valerdi, R. (2005). *The constructive systems engineering cost model (COSYSMO)* [Doctoral dissertation, University of Southern California, Los Angeles].

PANEL 20. MACHINE LEARNING: CHALLENGES AND OPPORTUNITIES FOR THE ACQUISITION WORKFORCE

Thursday, May 12, 2022	
11:05 a.m. – 12:20 p.m	**Chair: Hon Nick Guertin,** Director, Operational Testing and Evaluation, Office of the Secretary of Defense ***A System-of-Systems Approach to Enterprise Analytics Design: Acquisition Support in the Age of Machine Learning and Artificial Intelligence*** Cesare Guariniello, Purdue University Prajwal Balasubramani, Purdue University Daniel A. DeLaurentis, Purdue University ***Tips for CDRLs/Requirements when Acquiring/Developing AI-Enabled Systems*** Bruce Nagy, NAVAIR ***Recommending Recommendations to Support the Defense Acquisition Workforce*** Carlo Lipizzi, Stevens Institute of Technology Hojat Behrooz, Stevens Institute of Technology Michael Dressman, Stevens Institute of Technology Arya Guddemane Vishwakumar, Stevens Institute of Technology Kunal Batra, Stevens Institute of Technology

Hon Nick Guertin— was sworn in as Director, Operational Test and Evaluation on December 20, 2021. A Presidential appointee confirmed by the United States Senate, he serves as the senior advisor to the Secretary of Defense on operational and live fire test and evaluation of Department of Defense weapon systems.

Mr. Guertin has an extensive four-decade combined military and civilian career in submarine operations, ship construction and maintenance, development and testing of weapons, sensors, combat management products including the improvement of systems engineering, and defense acquisition. Most recently, he has performed applied research for government and academia in software-reliant and cyber-physical systems at Carnegie Mellon University's Software Engineering Institute.

Over his career, he has been in leadership of organizational transformation, improving competition, application of modular open system approaches, as well as prototyping and experimentation. He has also researched and published extensively on software-reliant system design, testing and acquisition. He received a BS in Mechanical Engineering from the University of Washington and an MBA from Bryant University. He is a retired Navy Reserve Engineering Duty Officer, was Defense Acquisition Workforce Improvement Act (DAWIA) certified in Program Management and Engineering and is also a registered Professional Engineer (Mechanical).

Mr. Guertin is involved with his community as an Assistant Scoutmaster and Merit Badge Counselor for two local Scouts BSA troops as well as being an avid amateur musician. He is a native of Connecticut and now resides in Virginia with his wife and twin children.

A System-of-Systems Approach to Enterprise Analytics Design: Acquisition Support in the Age of Machine Learning and Artificial Intelligence

Cesare Guariniello—is a Research Scientist at Purdue's School of Aeronautics and Astronautics. He received his PhD in Aeronautics and Astronautics from Purdue University in 2016 and his master's degrees in astronautical engineering and computer and automation engineering from the University of Rome "La Sapienza." Guariniello works as part of the Center for Integrated Systems in Aerospace and is currently engaged in research projects funded by the DoD, the National Science Foundation, and JPEO. His main research interests include modeling and analysis of complex systems and SoS architectures—with particular focus on space mission architectures, aerospace technologies, and robotics.

Prajwal Balasubramani—is a Graduate Research Assistant at Purdue's Center for Integrated Systems in Aerospace and is currently pursuing a PhD in aerospace engineering. He graduated with a BTech in aerospace engineering from Jain University, India, in 2018, and received a master's degree in Aeronautics and Astronautics at Purdue University in 2020. His primary research interests are in the areas of systems engineering, system dynamics, multi-agent systems, mechanism design, and digital engineering. He has also worked on various projects involving operations research, agent-based modeling, and system-of-systems.

Daniel A. DeLaurentis—is a Professor at Purdue University's School of Aeronautics and Astronautics. He leads Purdue's Center for Integrated Systems in Aerospace, and he is also the director of Purdue University's Institute for Global Security and Defense Innovation (i-GSDI). His primary research interests are in the areas of problem formulation, modeling, and robust system design and control methods for aerospace systems and systems-of-systems. This includes agent-based modeling, network theory, optimization, and aerospace vehicle modeling. His research is conducted under grants from NASA, the FAA, the Navy, the DoD Systems Engineering Research Center (SERC) UARC, and the Missile Defense Agency.

Abstract

System-of-Systems (SoS) capability emerges from the collaboration of multiple systems, which are acquired from independent organizations. Even though the systems contribute to and benefit from the larger SoS, the data analytics and decision-making about the independent system is rarely shared across the SoS stakeholders. The objective of the research presented in this paper is to identify how the sharing of datasets and the corresponding analytics among SoS stakeholders can lead to an improved SoS capability. Our objective is to characterize how appropriate use of data sets may lead to deployment of different predictive (predicting an outcome from data) and prescriptive (determining a preferred strategy) analytics and lead to better decision outcomes at the SoS level. We build and demonstrate a framework for this objective, based on extensive literature review, which utilizes appropriate predictive and prescriptive methodologies for SoS analysis. Additionally, we propose to utilize machine learning techniques to predict the achievable SoS capability and identify sources of uncertainty derived by sharing partial datasets. A case study demonstrates the use of the framework and prospects for future improvements.

Introduction

Acquisition in the context of System-of-Systems (SoS) presents additional challenges due to the independence of stakeholders, which is a characteristic trait of this category of complex systems. Data availability can be affected by uncertainty due to the independence of stakeholder decisions. Therefore, an approach is necessary which is suitable to understand how different SoS scenarios in acquisition can be addressed with appropriate strategies to minimize the risk due to uncertainty. The use of predictive analytics to model expected behavior of variables of interest, combined with prescriptive analytics which will support adequate decision-making in the presence of uncertainty, constitutes a first step to address the difficulties in SoS acquisition. However, it is often important not only to identify best practices to address specific scenarios, but to be able to assess patterns that characterize different types of problems. We therefore propose to utilize machine learning techniques to assess achievable SoS capability that can be achieved by sharing pertinent datasets and to prescribe the information links between systems to enable this sharing. This combination of predictive, prescriptive, and machine learning methods is the foundation to our acquisition support framework.

We use a case study to demonstrate the use of the framework and to identify future steps. Previous research used a Decision Support Framework (DSF), developed by researchers at Purdue University, to simulate and analyze a fictitious multi-domain battle scenario, where the different stakeholders do not agree on the relative weight of the different achievable SoS capabilities. This example did not make use of predictive and prescriptive methodology and addressed only uncertainty due to different stakeholder objectives. The case study in this work models an acquisition problem for an Urban Air Mobility service, where a stakeholder entering the market faces uncertainty in population commute data because of partial information on potential customers and competitor market strategies. Here we use predictive (regression modelling) and prescriptive analytics to provide support towards the decision-making and locally optimal acquisition, after properly modeling the interactions due to the dynamic nature of this SoS problem. This framework is then leveraged to conduct multiple experiments with varying scenarios for stakeholders to play out, in order to build a data set on which machine learning algorithms can be applied to extract key dependencies and factors in the market space. These insights then favor acquisition decisions to build an SoS.

Background of Research and Literature Review

Acquisition in a System-of-Systems Context

System-of-Systems (SoS) capability emerges from the collaboration of multiple systems, which are acquired from independent organizations. The systems within an SoS serve two purposes: one is to meet their own independent objectives, and the second is to contribute some capability to the SoS from which all constituents can benefit. In recent decades, the fields of machine learning and data analytics have found widespread application in system design and acquisitions. It is unanimously understood that any organization acquiring a complex system employs some form of data analytics to assess a system's independent objectives. Even though the systems contribute to and benefit from the larger SoS, the data analytics and decision-making about the independent system is rarely shared across the SoS stakeholders.

Characteristics of SoS (Maier, 1998) make them quite different from simple systems and the resulting behavior of a SoS is often unpredictable just by knowing its constituent parts, due to the interactions between those parts. Given the interdependencies in SoS, when considering acquisition, it is important to recognize the stakeholders, resources, operations, policies, and economics not only of one system, but of the entire SoS. Uncertainty and possible hidden information are common in SoS acquisition, and since the SoS capability is a multi-faceted enterprise, it is hard to formulate a single set of mathematical equations that would cover all cases. Therefore, in this work we develop research towards an information-centric framework that helps inform early-stage decisions on enterprise level.

Important context for our work comes from the ambitious goals put forth in both defense and commercial sectors for Digital Engineering (DE) and its related components in various engineering functions, such as Model-Based Systems Engineering (MBSE) for the SE domain. DE and MBSE pursue the use of digital models at every phase of acquisition. Within this context, the overarching goal of our framework is to examine the impact that data features (e.g., survey categories, types of variables, ownership/privacy of data, etc.) have on the type and effectiveness of predictive and prescriptive analytics that can be employed and how the outcome can be shaped differently by examining the connectivity of data sets. This is particularly important for SoS acquisition where these data sets exist at the local system level but may not be shared at the SoS/enterprise level or vice versa. Our objective is to characterize how the sharing and the connectivity of data sets may lead to deployment of different predictive and prescriptive analytics (due to data access) and lead to better outcomes at the SoS level.

Overview of Data Analytics

Predictive data analytics provides methodologies to anticipate and predict outcomes by collecting and utilizing prior information (Joseph & Johnson, 2013; Rehman et al., 2016; Waller & Fawcett, 2013). Although using data to guide decision-making has been around since the Babylonian times, where data was recorded on tablets to predict harvest (Lo & Hasanhodzic, 2010), a major shift in the ability to reason over large amount of data emerged in 1940s with the advent of computer development, storage, and machine learning techniques. For application in complex systems, early usage of analytics can be traced back to the 1940s and ’50s when data analytics models were used to predict outcomes for the behavior of nuclear chain reactions in the Manhattan Project and the weather forecasting using the ENIAC computer (Lynch, 2008).

Prescriptive data analytics, on the other hand, aims to provide an ability to generate/prescribe the best courses of action based on given information which may be obtained from a predictive data analytic outcome. Starting around World War II, the need to optimize courses of actions stimulated the development of operations research field, which in the proceeding decades led to *Analytics 1.0* for introducing data-based decision making in organizations. As the capabilities of computing and machine learning evolved to handle structured and unstructured large data sets (also known as Big Data), *Analytics 2.0* became the new paradigm across most large enterprises such as Google and Amazon (Davenport, 2013). Today, the Big Data landscape is shaped by the volume, variety, velocity, and veracity of data (known as the big four Vs of data science) and organization’s ability to include this *Analytics 3.0* in the decision-making process has become fundamental to its success and profitability. It will not be a generalization to state that most successful organizations employ some form of *Analytics 3.0* for business and product development.

For SoS acquisition and capability development, deployment of *Analytics 3.0* provides a unique challenge where the individual organizations contributing the constituent systems individually employ a suite of predictive and prescriptive analytics tools. However, these analytics and the underlying data sets are rarely shared across the SoS stakeholders. Given that the SoS capability emerges from the collaboration of otherwise independent systems and considering the ever-increasing need of interoperability between systems for transitioning towards DE and MBSE, there is an imperative to connect the data sets across SoS for holistic *Analytics 3.0* capability deployment. Previous work (summarized in the Machine Learning Techniques and Application in the DoD and First Steps from Previous Work: Optimal Acquisition with Uncertainty on Objectives sections) established the significance of utilizing Machine Learning techniques and predictive and prescriptive analytics to address uncertainty in SoS acquisition.

Machine Learning Techniques and Application in the DoD

This research builds upon previous work (Raz et al., 2020) which analyzed the use of Machine Learning techniques in DoD applications. Table 1 summarizes the findings from Raz et al., 2020, describing various Machine Learning methodologies, their assumptions, and applications in DoD research.

Table 25. Summary of ML Methodologies

Method	Key Features	Assumptions	DoD Reference
Supervised Learning			
Linear Regression	Fits quantitative/categorical predictors and continuous response to regression line using OLS	Linear parameters, constant error variance, independent error terms, errors are normally distributed, random sample of observations, no multi-collinearity	Moore and White III (2005)
Ridge Regression	Modification of linear regression that uses L2 norm when multi-collinearity assumption in linear regression is broken	Standardization of predictors, linear parameters, constant error variance, independent errors	Huang and Mintz (1990)
Lasso Regression	Used as a variable reduction or feature selection technique that shrinks some predictor coefficients to exactly zero to reduce overfitting from the linear regression model	Model has sparsity, irrepresentable conditions (Zhao & Yu, 2006)	Wang and Yang (2016)
Binary Logistic Regression	Models the log odds (using logit link) of a categorical binary outcome variable as a linear combination of quantitative/categorical predictors	Independent observations and errors, binomial distribution of response variable, linearity between logit of response and predictors	Apte et al. (2016)
Support Vector Machine	Uses a linearly separable hyperplane to classify data into two classes	Independent and identically distributed observations, margin is as large as possible, support vectors are most useful data points	Wei et al. (2006)
Artificial Neural Networks	Model consisting of interconnected nodes that receive inputs and return outputs based on an activation function	Independence of inputs	Brotherton and Johnson (2001)
K-Nearest Neighbors	Used to classify data points based on class that appears the most among neighboring points (classification) or average of classes (regression)	Similar inputs have similar outputs (Weinberger, 2018)	Xiao et al. (2006)
Naive Bayes Classifier	Uses Bayes theorem to calculate probabilities of a class response and selects the class with highest probability as the output	Predictors are conditionally independent of each other given the response	Freeman (2013)
Decision Tree	Algorithm that recursively and iteratively partitions the data into homogeneous subsets to identify a target outcome	Entire training set is at root node, quantitative predictors must be discretized	Apte et al. (2016)
Unsupervised Learning			
K-means	Use to identify homogeneous clusters in a data set	Clusters sizes are similar and spherical in form	Zainol et al. (2017)

Different stakeholders might use different Machine Learning techniques for prediction and decision making. In the presence of stakeholder independence, it is important to recognize what information is available and what information will be at least partially hidden (thus causing uncertainty), and then choose appropriate techniques to deal with the uncertainty. When the uncertainty is only on the desired objectives of stakeholders, we can design experiments to treat multiple cases and run predictions based on possible different choices of the stakeholders.

However, when the uncertainty is due to multiple, external factors (for example, competitors' decisions, stakeholder preferences, and fluctuation of the market, as shown in Figure 1), simpler predictive analytics are a better choice. Predictive analytics will provide baseline scenarios for subsequence application of prescriptive analytics, which can support educated decision-making that will cause robust decision based on the expected scenarios made available by the predictive analytics.

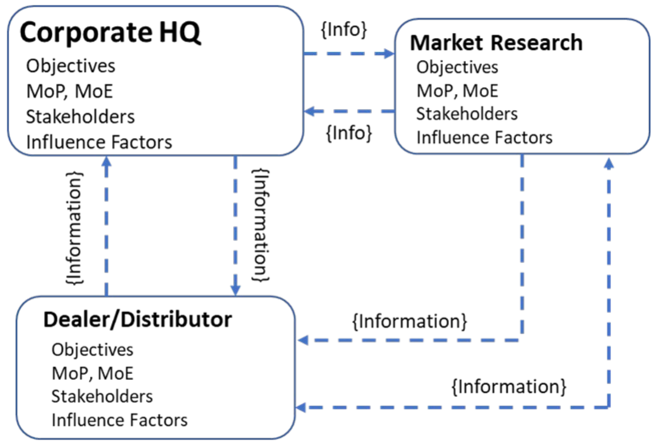


Figure 40. Conceptual Problem to Identify Impact of Data-Set Connectivity

First Steps from Previous Work: Optimal Acquisition with Uncertainty on Objectives

A precursor to this work, an application on a Naval Warfare Scenario (Raz et al., 2020) demonstrated the use of a Decision Support Framework (DSF) to assess optimal acquisition where multiple stakeholders might not agree on System-of-Systems-level objectives. The DSF identifies optimal portfolios of systems that, accounting for operations constraints, budget limitation, and uncertainty on capabilities, provide the best SoS performance.

Since in this case the uncertainty is due to different interpretation and preferences about mission requirements and objectives, the DSF has been simply used to run multiple scenarios, each one having a different combination of preferred SoS-level objectives. The resulting optimal portfolios of systems have then been compared to identify common occurrences of certain systems in optimal portfolios for any given budget limitation and risk acceptance. At the same time, these results identify parts where additional information might be required to support optimal decisions (for example, when the difference in preferred objectives results in extremely different optimal portfolios).

Figure 2 shows the different combinations of weights, representative of the importance given by different stakeholders to various SoS-level objectives in the Naval Warfare scenario. Figure 3 shows pareto fronts of optimal portfolios providing weighted SoS-level performance (vertical axis) based on budget (horizontal axis). The different lines represent cases from Figure 2, i.e., different stakeholder preferences and different weights given to SoS-level objectives.

Weights			
Cases	Air Superiority	Naval Superiority	Reconnaissance
1	0.8	0.1	0.1
2	0.7	0.2	0.1
3	0.7	0.1	0.2
4	0.6	0.2	0.2
5	0.6	0.3	0.1
6	0.6	0.1	0.3
7	0.5	0.1	0.4
8	0.5	0.2	0.3
9	0.5	0.3	0.2
10	0.5	0.4	0.1
11	0.4	0.5	0.1
12	0.4	0.4	0.2
13	0.4	0.3	0.3
14	0.4	0.2	0.4
15	0.4	0.1	0.5
16	0.3	0.6	0.1
17	0.3	0.5	0.2
18	0.3	0.4	0.3
19	0.3	0.3	0.4
20	0.3	0.2	0.5
21	0.3	0.1	0.6
22	0.2	0.7	0.1
23	0.2	0.6	0.2
24	0.2	0.5	0.3
25	0.2	0.4	0.4
26	0.2	0.3	0.5
27	0.2	0.2	0.6
28	0.2	0.1	0.7
29	0.1	0.1	0.8
30	0.1	0.8	0.1

Figure 2. Test Runs with Variation in Weight Distribution

Some stakeholder decisions result in higher SoS-level performance, and different budget levels cause different set of weights (and the resulting optimal portfolios) to provide better performance. We can therefore notice that any uncertainty in SoS capability preferences affects the resulting performance of the SoS portfolios. Since its point on the pareto frontier corresponds to a different portfolio of systems, we can compare the optimal portfolios to identify common systems and to assess areas where further information might be needed, if available.

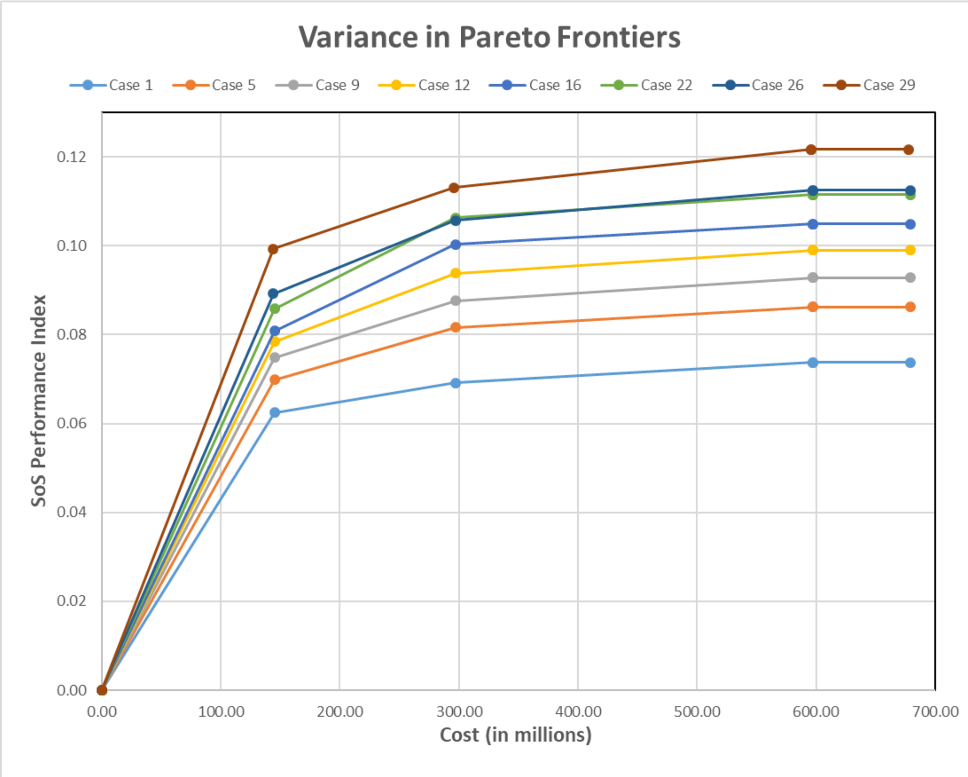


Figure 3. Variation in Pareto Frontiers Across the Cases

Optimal Acquisition with Uncertainty on External Factors

In this application, we consider much less "controllable" sources of uncertainty, and lack of information due to external factors. We model a scenario of Urban Air Mobility (UAM) in the Dallas, TX region, where a new stakeholder plans to participate as a provider of UAM services. The new provider has access (possibly limited) to data from the past that can suggest how many passengers would be willing to use UAM services. However, these data present some degree of uncertainty about the future, due to the many factors that can affect the number of passengers. Furthermore, the new stakeholder might not be fully aware of the decisions of competitors, which would affect the quota of available market to which the new stakeholder would have access. Predictive and prescriptive analytics, together with the use of Machine Learning, can support decision making in this context.

We model the expected user traffic in this transportation SoS by looking at previous years' data from the North Central Texas Council of Governments (NCTCOG). This is data on total population travel frequency between home and work districts, therefore it represents the total number of travelers between different locations. We then need to predict the proportion of travelers willing to use UAM services as an alternative to driving or using public ground transportation. Travelers' decision is affected by their income, the cost of commuting by UAM vehicles, and the perceived value of time. Figure 4, from Maheshwari et al. (2021) shows the process of comparison of UAM routes against ground transportation, when there are no potential competitors. This comparison gets specialized to each specific region of interest.

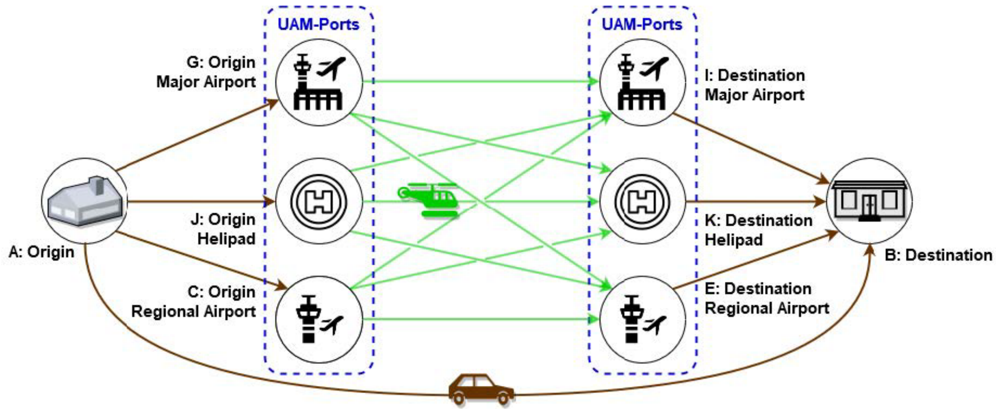


Figure 4. Alternative Routes of Travel from A to B: UAM Vehicles with Different Origins and Destinations and Ground Transportation

Urban Air Mobility Scenario and Problem Setup

Three regions in the Dallas-Fort Worth area (called A, B, and C in the results) are modeled in our study as potential UAM points of operation based on the NCTCOG data on population travel frequency. Figure 5 shows the location of vertiports for UAM in this study and the density of origins and destinations pertaining to travelers who would prefer UAM transportation. We model 6 routes operating between these three regions. Additionally, the stakeholder is provided with acquisition decisions for the type of UAM vehicles with varying seating capacity (1, 2, and 4 maximum passengers). These vehicles have different operational costs and different ticket price. The stakeholder decisions to build their UAM portfolio are motivated to maximize the expected total income per day. Our predictive model uses historical data from NCTCOG to predict the population travel frequency for 2022 for which the UAM network is being modelled as the acquisition problem. Since the NCTCOG data is quite sparse, we used simple linear regression models. We run an optimization problem with constraints on the maximum number of allowed flights per day and the maximum licensed number of vehicles for UAM. The decision variables are the number of acquired vehicles per type (seating capacity) per route. We extend the optimization to include the uncertainty in the predicted data, after running the predictive analysis with a 95% confidence interval.

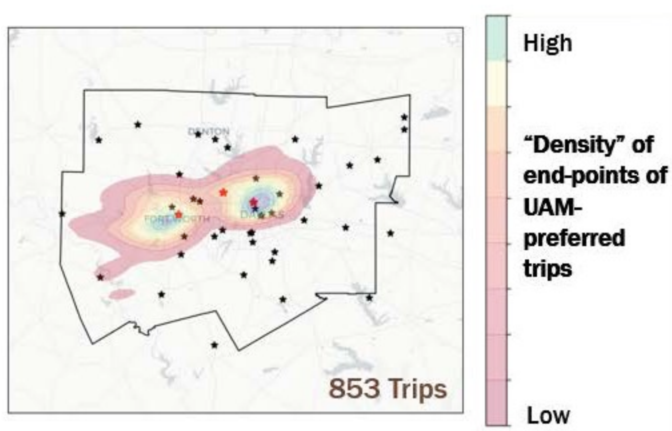


Figure 5. Location of Vertiports (Red Stars) in the Dallas–Fort Worth Area and Areas of Preference for UAM Transportation

Within this context, we also consider that UAM operators might be carrying on their activities in a competitive market, where other stakeholders are present. Based on the amount of available data concerning the travelers' decision (in its turn, influenced by income, cost of commuting by UAM vehicles and perceived value of time), the presence of competition, and some degree of uncertainty, we again compute the expected optimal choices for UAM operators to compete in the market.

Results—Predictive Analytics and Optimization

In the first example, we assumed perfect information about the number of passengers willing to fly with UAM vehicles, based on their income, the alternative ground-based travel time, and the personal perception of the value of time. The only source of uncertainty in this case is due to the prediction of travel frequency for 2022, based on past travel data starting in the year 2010. We also assumed perfect knowledge about the share of the market which is already taken by existing competitors. The income is based on ticket price (varying by route and type of vehicle) and on operational costs. We ran an integer linear optimization, where the decision variables are the number of vehicles per type per route, and the number of passengers actually flying with the stakeholder's vehicles. Constraints exist on the maximum number of vehicles that can be acquired and on the maximum number of flights per vehicle per day on each route. Results of this optimization based on the expected values for flying passengers in 2022 are shown in Table 2.

Table 2. Optimal Choices with Full Data Available and Prediction for the Year 2022 in the Dallas Area

Route	**1-passenger vehicles**	**2-passenger vehicles**	**4-passenger vehicles**
AB–BA	209	0	83
AC–CA	31	160	0
BC–CB	0	0	157
Passengers per day	2522	3840	11448
Income	$ 1,954,910.73		

Results show that on two routes with more passenger availability, larger vehicles are preferable even if they produce less income per passenger. The intermediate-sized vehicle is present only on two routes (from A to C and from C to A), together with the small vehicle. The expected income is about $1.955 million. However, due to uncertainty, the actual income will be slightly smaller. We ran 1000 scenarios according to the expected distribution and using the optimal choice of vehicles, which resulted in an income of $1.893 million.

As a first step towards the study of support for decision-making in SoS, we then ran a scenario where the actual market share is unknown. We assumed that 33% of the passengers are available to fly with the new stakeholder, which is only slightly different from the actual market share that we used in the first case (ranging between 30% and 45%). We also increased the uncertainty in the predictive phase. Despite the small differences, there are already changes in the choice of optimal fleet, as shown in Table 3. The routes AB and BA saw an increase in the number of small vehicles, while more large vehicles were acquired for the routes BC and CB.

Table 3. Optimal Choices with Partial Data Available and Prediction for the Year 2022 in the Dallas Area

Route	1-passenger vehicles	2-passenger vehicles	4-passenger vehicles
AB–BA	218	0	67
AC–CA	22	160	0
BC–CB	0	0	173
Passengers per day	2488	3840	11832
Expected income	$ 1,956,211.60		

The expected larger market share along the routes BC and CB suggests an income slightly larger than the previous case. However, despite the very small differences, the presence of incomplete data causes suboptimal acquisition. When running 1000 scenarios according to the expected distribution of flying passengers, the resulting income is $1.858 million, lower than the income achieved with the optimal choice in the previous case and about 5% lower than the expected income. These results show how even just a small amount of uncertainty can have a large impact on the decision-making process and its outcome.

Machine Learning to Enhance Prescriptive Analytics

To further extend the stochastic optimization, it can be useful not only to know the results of optimization in different scenarios, but also to understand how the different inputs (which are the source of uncertainty) affect the output variables. We therefore trained a Neural Network, implementing 1090 scenarios with variable parameters, which modify the optimization problem. For each route and type of vehicles, the parameters include the maximum number of flights per day, the market fraction available to the new stakeholder, and the feasible gain margin (that is, how much the stakeholder is desiring to earn out of selling tickets. This needs to overcome the operational cost, but high prices of tickets will cause fewer travelers to choose UAM vehicles over ground transportation).

Figure 6 shows a neural network trained in Matlab, where the inputs are different level of maximum number of flights per day, market fraction available, and feasible gain margin on routes AB and BA for vehicles with 1 and 2 passengers. The output variables are the result of the optimization with full data available and prediction for the flying passengers in 2022.

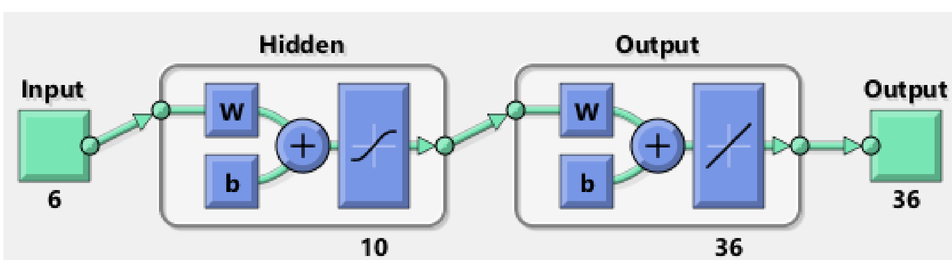


Figure 6. Neural Network for the UAM Scenario in the Dallas–Fort Worth Area

Figure 7 shows the result of the training of this partial Neural Network. Outliers are caused by the fact that this network is based on a partial number of inputs. However, the fit is good enough to utilize the network for prescriptive analytics and to quickly run analysis of a large number of scenarios with changing inputs. For example, Tables 4 and 5 show the outcome of the Neural Network with different inputs for the parameters. When variables pertaining to market fraction and maximum number of allowed flights per day increase for 1-passenger vehicles and decrease for 2-passenger vehicles, we can notice not only how this impacts the route directly affected by the parameters (AB and BA in this case), but also how the variables on available market share and desired gain margin, united with the number of passengers which will decide to use UAM vehicles, affects the choice of acquiring more 4-passenger vehicles. Limitations of this approach and solutions to overcome the limitations are presented in the following section.

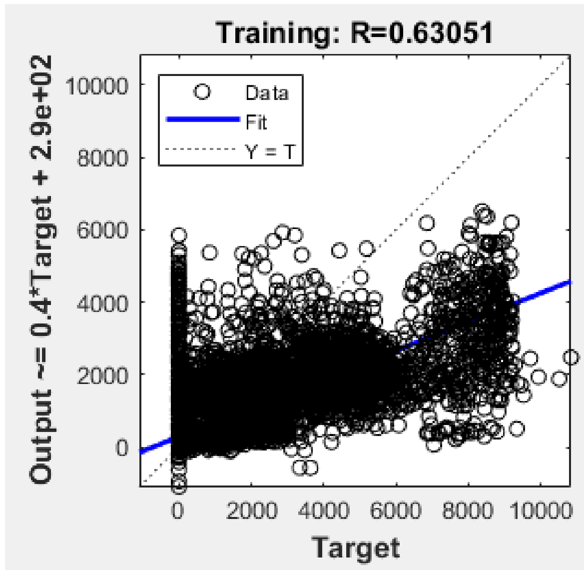


Figure 7. Fit of Training Runs of the Reduced Neural Network for the UAM Scenario

Table 4. Output of the Neural Network with Similar Values for Max Number of Flights Per Day, Market Fraction, and Gain Margin for 1-Passenger Vehicles and 2-Passenger Vehicles on Routes AB and BA

Route	1-passenger vehicles	2-passenger vehicles	4-passenger vehicles
AB–BA	**245**	**151**	129
AC–CA	22	94	144
BC–CB	2	41	105

Table 5. Output of the Neural Network with Higher Max Number of Flights Per Day, Market Fraction, and Gain Margin for 1-Passenger Vehicles with Respect to 2-Passenger Vehicles on Routes AB and BA

Route	1-passenger vehicles	2-passenger vehicles	4-passenger vehicles
AB–BA	**274**	**125**	179
AC–CA	24	159	221
BC–CB	2	45	118

Conclusions and Future Work

Building on top of previous research on the use of predictive and prescriptive analytics for acquisition in a System-of-Systems context, we expanded our framework that deals with the uncertainty derived from potential lack of data and information, to treat cases where the uncertainty is due to external factors and can heavily affect the outcome of decisions. The presence of incomplete data, together with uncertainty due to fluctuation of variables in the future and with the presence of independent stakeholders, can produce suboptimal choices in acquisition. Preliminary results showed promising directions for the use of predictive and prescriptive analytics to address this type of problems. An application to an Urban Air Mobility scenario in the Dallas–Fort Worth region was used for various demonstration purposes. First, we showed how decision-making in the presence of full information about some variables (fraction of market availability) while others are affected by limited uncertainty (prediction models for number of passengers willing to use UAM services based on income and perceived value of time) produce results very close to a global optimum in the choice of UAM vehicles to acquire. On the other side, even small changes in the availability of data about market distribution can cause suboptimal decisions. To alleviate the impact of uncertainty and to be able to analyze many scenarios, so as to support prescriptive analytics for decision-making, we propose the use of Neural Networks, that can be trained to provide insight into the dependency of variables of interest (in this case, acquisition decisions) on multiple inputs (in this case, for example, desired gain margin, available market fraction, and maximum allowed number of flights per day). This use of various Machine Learning techniques provides a first step into understanding the reasons for observed outcome, and therefore a step towards robust decision-making in the presence of uncertainty in a SoS. However, we also propose various refinements for future work. First, in this example we trained a Neural Network with a subset of the inputs. The Neural Network is trained as a whole and, other than implementing some basic regularization, does not quantify the impact that each input variable has on the outputs and the variability of these input-output relationships. Therefore, the use of Uncertainty Quantification can greatly improve the approach to these problems, by providing quantitative measurements of the importance that each input variable has on the outputs. This, in turn, will provide information on critical areas where more information (or more caution) is needed. Further research also includes extension of the case studies, to include non-recurrent and recurrent costs, and to use stochastic optimization techniques as additional prescriptive methodologies in support of acquisition decision-making.

References

Apte, U., Rendon, R., & Dixon, M. (2016). *Big data analysis of contractor performance information for services acquisition in DoD: A proof of concept*. Naval Postgraduate School Monterey United States.

Brotherton, T., & Johnson, T. (2001, March). Anomaly detection for advanced military aircraft using neural networks. In *2001 IEEE Aerospace Conference Proceedings (Cat. No. 01TH8542)*, 6, 3113–3123.

Davenport, T. H. (2013). Analytics 3.0. *Harvard Business Review*, *91*(12), 64–72.

Freeman, C. E. (2013). *Multivariate and Naïve Bayes text classification approach to cost growth risk in Department of Defense acquisition programs*.

Huang, C., & Mintz, A. (1990). Ridge regression analysis of the defence-growth tradeoff in the United States. *Defence and Peace Economics*, *2*(1), 29–37.

Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. *It Professional*, *15*(6), 43–48.

Lo, A. W., & Hasanhodzic, J. (2010). *The evolution of technical analysis: financial prediction from Babylonian tablets to Bloomberg terminals* (Vol. 96). John Wiley & Sons.

Lynch, P. (2008). The ENIAC forecasts: A re-creation. *Bulletin of the American Meteorological Society*, *89*(1), 45–56.

Maheshwari, A., Sells, B. E., Harrington, S., DeLaurentis, D., & Crossley, W. (2021). Evaluating impact of operational limits by estimating potential uam trips in an urban area. In *AIAA Aviation 2021 Forum,* 3174.

Maier, M. W. (1998). Architecting principles for systems-of-systems. *Systems Engineering: The Journal of the International Council on Systems Engineering*, *1*(4), 267–284.

Moore, G. W., & White III, E. D. (2005). A regression approach for estimating procurement cost. *Journal of Public Procurement*.

Raz, A., Balasubramani, P., Harrington, S., Guariniello, C., & DeLaurentis, D. A. (2020). *System-of-systems acquisition analytics using machine learning techniques*.

Rehman, M. H. ur, Chang, V., Batool, A., & Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management*, *36*(6), 917–928.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, *34*(2), 77–84.

Wang, R., & Yang, W. (2016). Estimation of demand and supply in U.S. airline industry. *Unpublished Working Paper*.

Wei, K., Wu, J., Ma, W., & Li, H. (2019). State of charge prediction for UAVs based on support vector machine. *The Journal of Engineering*,*2019*(23), 9133–9136.

Xiao, H., Cai, C., & Chen, Y. (2006). Military vehicle classification via acoustic and seismic signals using statistical learning methods. *International Journal of Modern Physics C*,*17*(02), 197–212.

Zainol, Z., Azahari, A. M., Wani, S., Marzukhi, S., Nohuddin, P. N., & Zakaria, O. (2018). Visualizing military explicit knowledge using document clustering techniques. *International Journal of Academic Research in Business and Social Sciences*, *8*(6), 1127–1143.

Zhao, P., & Yu, B. (2006). On model selection consistency of Lasso. *The Journal of Machine Learning Research*, *7*, 2541–2563.

Tips for CDRLs/Requirements when Acquiring/Developing AI-Enabled Systems

Bruce Nagy—is an applied Research Engineer at the Naval Air Warfare Center, Weapons Division (NAWCWD) at China Lake California. His focus is on integrating game theory and machine learning to create recommendation engines that handle complex, highly dimensional battle management scenarios. He also uses causal learning techniques to provide explainable, statistically based solutions to wargaming stratagem. When he served in the Navy, Nagy qualified as an Engineering Duty Officer. In that capacity, Nagy developed statistically-based algorithms to enhance satellite communications. He was also a technical troubleshooter for NRO and received honors for his efforts in recovering a failing national defense program that jeopardized U.S. global security. In addition to his wargaming and machine learning research, he is driving a joint effort to create standards that establish an appropriate level of rigor for the acquisition and/or development of AI-enabled systems for deployment. He has received four degrees in math, science and engineering, and has done postgraduate work analyzing brainstem to muscle group neurology. [bruce.m.nagy.civ@us.navy.mil]

Abstract

The Department of Defense (DoD) is challenging Acquisition professionals to manage the development of systems incorporating AI functions either as upgrades or new programs of record. But, AI functions present unique challenges associated with requirements and subsequently when creating a suitable Contract Data Requirements List (CDRL). The problem stems from the ability to ensure the quality and quantity of training data sets which can limit the reliability of AI performance. Currently, there is limited guidance regarding topics for discussion during an AI requirements review or as to what AI related information should be required in CDRLs. However, a recent investigation into the lack of AI development guidelines prompted a NOSSA-funded project. Using an AI "sandbox" approach, a DoD representative program, involving AI/ML algorithms supporting a mission planner with autonomous vehicle selection and navigation, was used to determine realistic requirements specific to systems incorporating one or more AI functions. As a result of their analysis, this paper presents contents for an AI Development Plan (AIDP) to be part of a CDRL. Within the AIDP, measurements and new evaluation methods are also offered, as well as questions and considerations to support quality AI development.

Common AI Acquisition and Development Issues

Issues facing the acquisition professional managing a program implementing AI are: (1) heavy use of AI jargon where folks among the data science community can't seem to agree on common definitions, (2) a lack of understanding of the core workings of various AI algorithms and thereby turning a function into a seemingly magical black box. However, once the vocabulary is understood, I believe the acquisition professional is much further ahead. The core workings of an algorithm require only an understanding of the mechanics without needing to go into the detailed math, i.e., system engineering constructs in terms of DoD Architecture Framework (DoDAF), Unified Modeling Language (UML), etc.—those are things that can be left to AI software engineers. And finally (3) having confidence that the training data accurately represents what the AI algorithm will experience during "stressful" deployment periods.

The first problem, less technical and more cooperative, can be solved by the program sponsor and the prime contractor agreeing upon a common AI dictionary of terms. The dictionary can be created by either the DoD, commercial sector or a self-created source. The second problem, educating program participants in the basics of AI, is also a potentially solvable issue through training. The third problem is the Achilles heel for anyone acquiring an AI-based system. There is currently no "official" guidance on how someone involved in AI development should view training data in terms of quality and quantity. Does it need to be created from "live" operational environments or will M&S synthetic data be sufficient? Current work in this area demonstrates that AI training data issues cause a significant amount of redo work, cost overruns, and schedule delays. M&S synthetic data has both pros and cons associated with its use as training data. For the acquisition professional, knowing how to address these issues is absolutely necessary. Because of the deficiency gap in AI policies, guidelines, and tools, acquisition professionals, from PM, SE, and T&E, as well as system safety practitioners need education/training regarding how to assess AI development to ensure confidence in the behavior of current WD weapons programs implementing AI upgrades. Because of these deficiencies and by direction of DoD, many working groups are attempting to understand AI's unique software requirements, architecture, design, code, and test without any reference as to what works and what does not, instead are using ad-hoc approaches. This is dangerous!

Background to Determining Documentation Content

Naval Ordnance Safety and Security Activity (NOSSA) funded this research over a two year period to investigate unique AI/ML policies, guidelines, tools and techniques to assess safety in identified critical functions. The project, as shown in Figure 1, involved two autonomous robots delivering packages, using an intelligent route planning system that considered the degree of difficulty with the routes, including crime, weather conditions, and human factors. The sandbox approach was used to mock-up an AI development process for the purpose of creating AI system of systems analysis to examine process and requirements. From this work, documentation structures became evident.

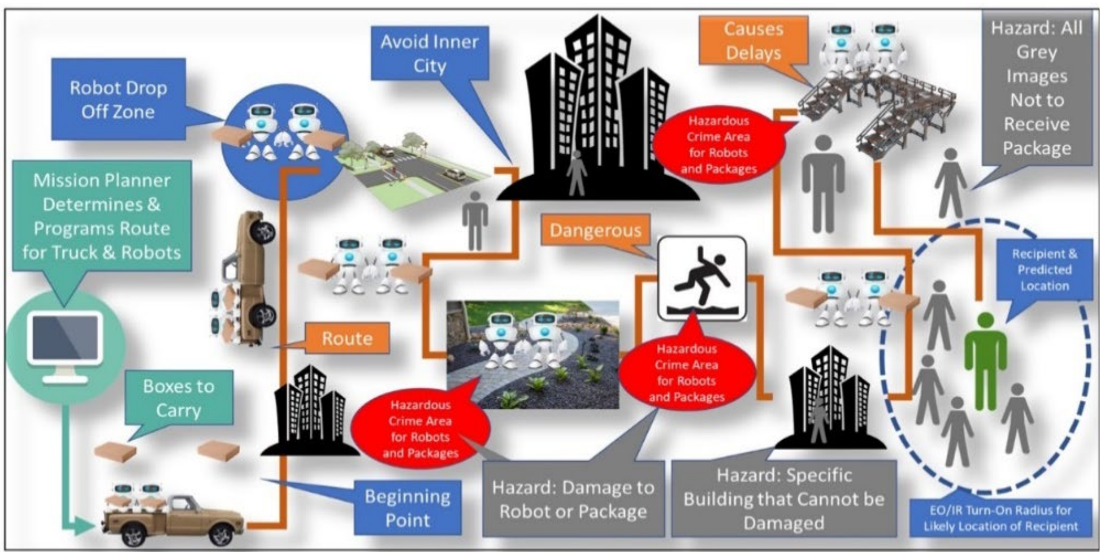


Figure 1. Operational Use Case

From this sandbox approach: (1) DoDAF and UML diagrams were created that identified MSG, API and SQL protocol requirements, (2) detailed architecture and designs were reviewed, and (3) software code was written, including a simulation program that modeled the use case described above—this allowed for training data to be acquired consisting of five classes and 17 attributes. Using the sandbox development approach allowed for the creation of level of rigor (LOR) tasks appropriate for each of the five stages. Through this analysis and development process, LOR was created in the form of practical questions and considerations to rigorously ensure AI/ML systems are built to have confidence in their functional behavior associated with the five stages: (1) Requirements, (2) Architecture, (3) Algorithm Design, (4) Algorithm Coding, and (5) Test and Evaluation. This paper focused on the requirements portion of this research as described in various documentation formats.

Contract Data Requirements List (CDRL)

A common CDRL approach is to list a Software Development Plan (SDP). However, SDP's have a wide range of content formats without a common structure. The DoD attempted to create a standard but found it difficult. The closest standard as guidelines can be found in DOD-STD-7935A, *Military Standard: DoD Automated Information Systems (AIS) Documentation Standards* (1988). "These standards provide guidelines for the development and revision of the documentation for an automated information system (AIS) or applications software, and specify the content of each of the 11 types of documents that may be produced during the life cycle of an AIS" (DOD-STD-7935A, 1988).

Many standards have come and gone, e.g., MIL-STD-498, DOD-STD-2167A, and DOD-STD-1703, due to lack of flexibility. Issues like a standard focus on waterfall project management styles, versus Agile type development, or not using modern software development tools, such as Computer-Aided Software Engineering (CASE) tools.

The SDP goal is to determine the scope of a software development effort, with guidelines in place to ensure quality development. However, based on history with regard to software, it's been difficult to create a one size fits all solution. A solution to this approach is in using MIL-STD-31000 *Technical Data Packages* (TDPs), which allows for customized content development. Another approach is in using DI-MISC-80508B, Data Item Description: Technical Report-Study/Services 2006), which also allows for customization.

Due to this challenge, this paper recommends a separate document, an AI Development Plan (AIDP), learning from common practices involved with customized SDPs, TDPs, and Technical Report Studies. The AIDP focused on AI process and guidelines that follow a level of rigor approach derived from the previously mentioned sandbox research. Therefore, the suggestions are in the form of questions and considerations listed specifically to the topic associated with various sections. The questions and considerations support an increased confidence in the behavior of an AI system when deployed.

Tips for CDRL Documentation

Five types of AI focused documents will be discussed with an emphasis on content: (1) AI Justification Report—3 Sections, (2) Best Practices Report—2 Sections, (3) Measurement Report—3 Sections, (4) Test and Evaluation Readiness Report, and (5) Missing and Sparse Class Tables—4 Sections. Each document type has a specific purpose as will be discussed. These five types of documents comprise the AIDP.

AI Justification Report

Section 1—AI identification: Conduct AI Type Function analysis of the proposed functions to determine if it includes an AI algorithm. If it is identified as an AI Type Function, then the document types and related sections contained in this report are suggested.

The hypothesis is that a function is an AI Type Function candidate if one or both criteria are met:

- (Criteria 1—Data Approximations) The function requires the use of <u>*data approximations*</u> to build/train its algorithm. Approximations can sometimes lead to inaccuracies. For example, it might use a single value, like average speed. An algorithm might use average speed to determine a time instead of the actual speeds encountered during deployment. This could lead to a decision error. Another example of data approximation which could also lead to decision errors is the use of text to represent values. For instance, representing many altitudes by the word "high." In this case, the data approximation "high" represents an infinite number of altitude values above a certain threshold. The function may be designed to make a decision when the data input states "high," potentially causing a decision error. Simulations use data approximation to model dynamics. AI algorithms might use synthetic training data sets, again potentially causing algorithm performance errors. A common approximation data inaccuracy concern is when training sets are synthetically generated. The concern is whether the data generated is replicating "realistic" background noise. This type of data approximation inaccuracy found in synthetic data has been noted as a main cause of an AI algorithm's poor performance when deployed. In game theory, the payoff tables are normally based on mathematical approximations called expected utility functions (EUF). If an algorithm uses a payoff table to make a decision, the EUF approximation might cause decision errors.

- (Criteria 2—Data Samples) The function requires the use of *data samples* to build/design its algorithm. For training, data normally consists of a representative subset of all the data contained in a larger population. Sample representations of the larger population can sometimes lead to inaccuracies. Subsets can be created from "live" or synthetically generated (simulation) sources. If the function requires the use of synthetic data that created the samples, the concern would be how much of the model approximated "reality" and how much of the total population was synthetically covered within the created subset? An example of using samples from "live data sources" would be snapshots of images describing facial expressions. The concern is whether the collected facial expressions within the training set represent all images that a person might express over a period of time when deployed in a variety of situations. If the subset of images collected of facial expressions do not adequately represent the deployed experience, then the training of the algorithm is limited. Will this limitation caused by the subset cause the algorithm to have performance issues? Another challenge associated with creating adequate data sample representation is when collecting information from experts or other authorities to create a decision tree. Decision trees requires input and output rules. Normally only a subset of all possible input and output combinations are used. Again, because only a subset is used, how well this subset represents deployed input and output challenges will determine the algorithm's performance.

Section 2—AI Scope: Discuss and document a justification for the proposed use of the AI algorithm vs. using a more traditional software, firmware or hardware technique. The goal is to explain why an AI algorithm is a "better" fit to the functional requirement.

Verify that an AI/ML function is needed by asking the following questions:

- For Criteria 1, data approximations: Can the algorithm be traditionally built using data approximations? Why or why not? A question to consider is, "Could another developer create a different set of statistics under the same conditions?" If no, then maybe this algorithm is not an AI Type Function. If yes, then there are data approximations and determining how one approximation is better than another needs to be considered. As an example, if a statistical model of the function was developed, how accurate were the approximations used in creating the function. In other words, how close do these approximations fit the physics of the real-world regarding operational deployment? How accurate is the distribution? Another consideration is investigating if the data approximation can be decomposed into greater detail, thereby reducing the size of the approximation. The goal is to have accurate data approximations that will result in quality training sets.
- For Criteria 2, data samples: Can algorithm be broken into subpopulations to allow development of traditional code? Why or why not? Another question to consider is, "What is the actual population size of the training set?" If the training set is equal to the

actual population size, then the function does not need an AI approach and can be handled traditionally. Consider the most basic ML algorithm, a regression line. If all the points that will ever occur for this function are used on the scatter plot-to approximate the curve, why use a regression line? If all the ML algorithm inputs and outputs are known, why use ML and not traditional code, i.e., if this, then that? Again, if traditional code can address the needs of the function, then that should be the approach used. If the function is based on simulation results creating data samples, then the concern is the "garbage in, garbage out" issue—poor real world representative synthetic data will result in an inferior model. The goal is to have comprehensive data samples that will result in robust training sets.

Section 3—AI Autonomy: Discuss and document a justification for the AI algorithm's level of autonomy, i.e., lack of supervision. The goal is to explain why an AI/ML algorithm requires the selected level of autonomy based on functional performance requirements and not a lower level.

a) Document how the design can or cannot include human in the loop oversight or traditional hardware/software technology acting as a guiderail/guard to provide checks and balances.

b) If checks and balances are limited, provide documentation as to operational limitations by:

1. Describe weaknesses of each AI/ML technique, e.g., expected success rate of the function. For example, if AI/ML is built on data approximations (using AI Type definition), how much bias will the data approximations add to the functional outcome? Or, if AI/ML is built on data samples, how representative are the samples to the population?
2. Determine how the training data is being generated, e.g., truth, synthetic, combination. Are these sources valid? Why?
3. Where is the training data coming from? Is it enough? (Remember the more sophisticated the AI/ML software, the more likely that it needs larger amounts of training data)
4. Will an outside independent source review the training, validation and test data created? Why or why not?
5. Will an outside independent source validate the success rate of the AI/ML function as compared to other AI/ML functions used in industry? Why or why not?

Best Practices Report

Section 1: Modality Type

What Machine Learning Training Data modality type are you representing in your deployed system and your data generation process? When creating training data, it is important to understand the operational environment being represented in order to ensure adequate development of the machine learning (ML) algorithms. The training data is either found from live events or synthetically created to match the operational scenario that will be provided as input to the ML algorithm. Therefore, the ML algorithm must learn how to perform under these conditions. Three types of modality represent various operational environments that can be encountered during deployment, where the type of modality defines how the ML algorithm needs to be trained. Understanding ML training data modality is fundamental to developing a reliable AI system (Nagy, 2021).

ML Training Data Modality 1: This modality, shown if Figure 2, supports training data sets that are based on an operational environment from multiple data sources, where each source contains one or more attributes. The various sources of separate data attributes are either found from live events or synthetic simulations created to match the deployed operational scenario. Therefore, the input for the ML algorithm for training needs to replicate the input that will be received during deployment.

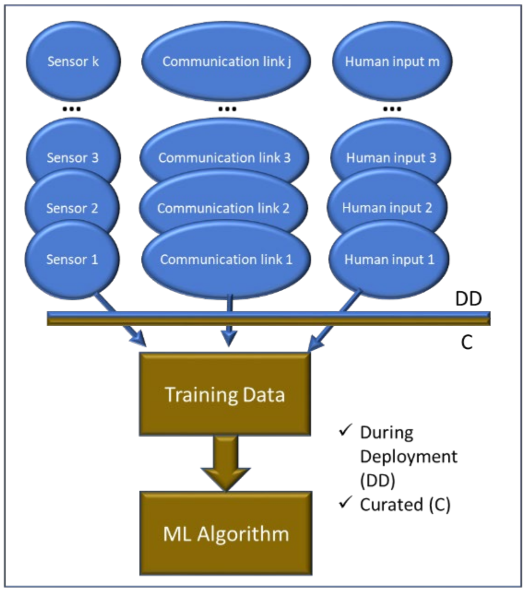


Figure 2. Example of Modality 1 Receiving Attribute Values from Various Sources

ML Training Data Modality 2: Training data sets that are based on an operational environment from a single data source, where the single data source contains multiple data attributes, as shown in Figure 3. The one stream set of aggregated attributes is either found from live events or synthetic simulations created to match the deployed operational scenario. Therefore, the input for the ML algorithm for training needs to replicate the input during deployment.

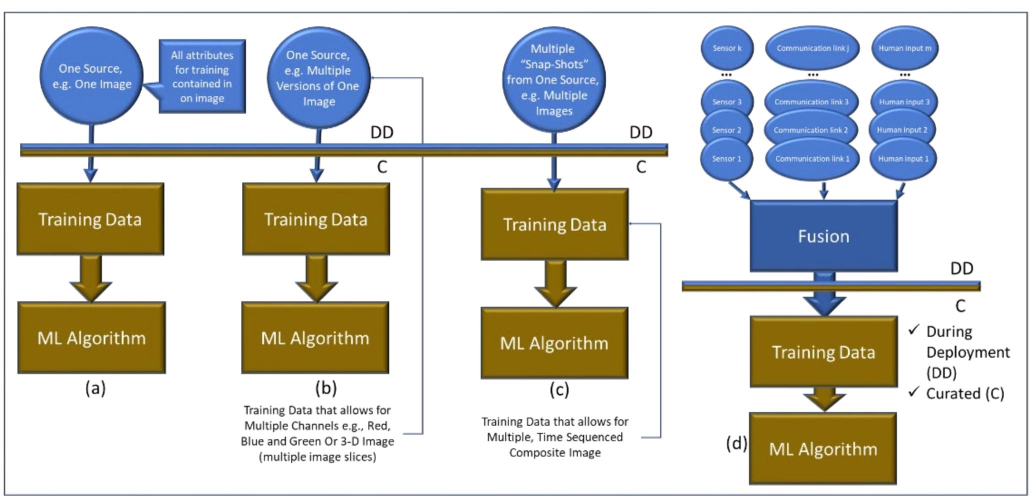


Figure 3. Examples of Modality 2 Training Requiring Images (a, b and c) or Fused Attribute Data (d)

ML Training Data Modality 3: Training data sets that are based on an operational environment from a combination of multiple data sources, shown in Figure 4; each source contains one or more attributes from various sources and from a single source containing multiple aggregated data attributes. It is a combination of Modality 1 and 2 that the algorithm uses for categorization or regression.

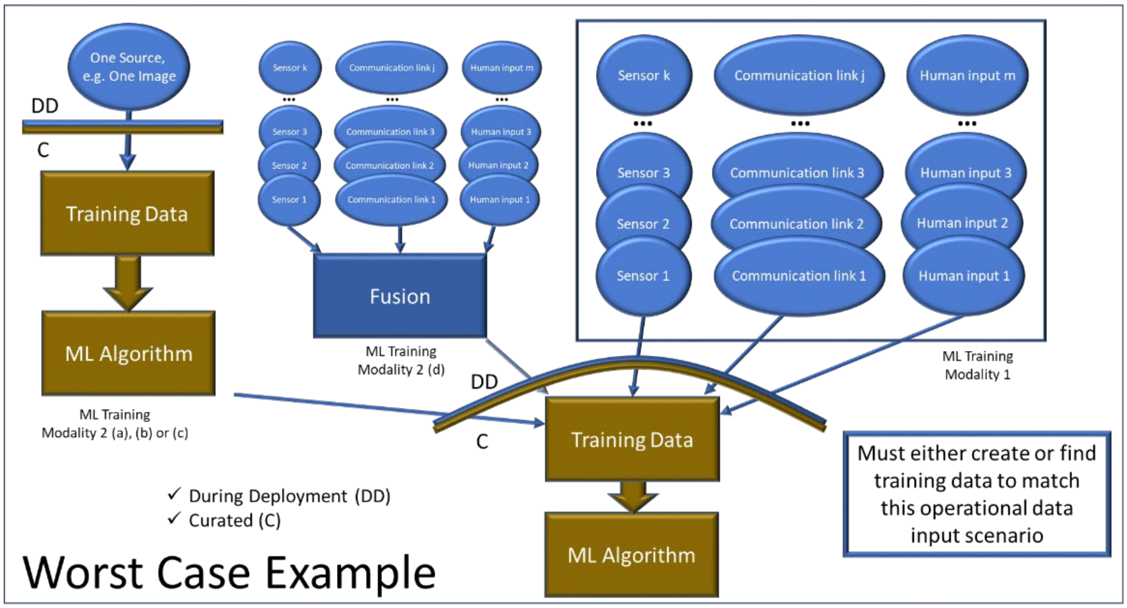


Figure 4. Example of Modality 3, a Combination of Modality 1 and 2

Instances/samples comprising training sets are composed of a combination of attributes, sometimes called features. When a feature is identified within an image, it is described as a piece of information contained in the content of the image. In this case, the feature describes a specific region of the image, which has certain properties, as opposed to another popular definition of a feature, a single pixel in an image.

The aggregation of attributes can be contained in one source, e.g., a camera taking a facial picture, or from many sources, e.g., various sensor inputs, such as radar and communication links. In this paper, we will distinguish whether attributes are generated from one or multiple sources based on their modality.

Adversarial ML is not considered a system safety issue, but does affect AI model confidence. It is important to know that it introduces challenges in the behavior of an AI model. Adversarial ML is modality dependent. Adversarial ML is most-times designed to cause an error in the output of the AI model being targeted.

Based on modality design, how can the deployed AI model be exploited by an adversary? Is this a consideration in Operational and Systems Requirements Discussions/Reviews in terms of the behavioral confidence of the algorithm or training set adequacy?

If Adversarial Networks/attacks become a consideration for any of the AI Functions under review, then an analysis by the developer to support the concern should be provided that includes describing how the balance between quality and vulnerability of the training set is being achieved with some form of objective, measured precision.

Section 2: Dataset Structure

Part 1—Representation: Does the synthetic or live data represent all the training data needed to train the algorithm to identify each label/class within the needed success rate? Examples of classes are various types of targets, described by a label, that are determined from the output of the algorithm. Using data sets to train the algorithm to identify an object is a typical ML process.

If not, how are classes being represented; are values being determined by using ML for regression? This question relates to how classification or regression is accomplished. For classification, algorithms can have two to many classes. For regression, then some form of analysis is used to determine a number or range of numbers. It's important to understand how the algorithm needs to perform and what type of data is being used to train an algorithm to perform adequately.

Note: A class, also referred to as target, label, or category, is what a categorization algorithm labels an input. For instance, if only an image of a cat or dog is used as input/training data for an ML algorithm, then the algorithm only has two classes either a cat or dog. To make a determination, each class would normally have a threshold value that would have to be met. If that threshold value is not achieved, by either class, then the AI model would fail to determine the input. For example, if the input was a coffee cup instead of a cat or dog, the threshold value would not be sufficient for either class and the input would not be determined.

Part 2—Fitness of the Data: Does each class have an appropriate number of attributes, or values that can be learned by the algorithm for the class/number being determined? In other words, has overfitting and underfitting been considered for each class/number with regard to the quantity of attributes/values simulated/collected and does that quantity reflect real world operations?

- Overfit. Indicates that there is an issue with the quality of the quantity of training data used. Overfit occurs when the algorithm's success is limited to a small amount of input variations when compared to the original training data. Limited input variations mean that as long as the input instance/sample closely matches an instance/sample of the training data, then it will accurately categorize/approximate, otherwise the algorithm will likely generate an error. Overfit states that it has a very low tolerance for input data that is not close enough to the original training data input.
- Underfit. Indicates that there is a quantity issue with the training data used. Underfit occurs when there is an insufficient amount of training data which causes missed categorizing of the correct class/approximation.

Both indicate a poor level of performance. Therefore, how will it be determined that the training data, attributes, or values within the instances used for categorization/regression, be sufficient to maximize success for data inputs not in the original training set? Note: This discussion must evolve around how the algorithm will be operationally deployed.

Part 3—Mission Alignment: How do we know that the synthetic or live data creating the training data is aligned with the mission parameters?

- Was a traceability study performed to support adequate coverage?
- Have statistics been shown on the number of configurations available and the number trained using data sources?
- How are we avoiding overfitting and underfitting based on training mixes and sets?

- Is the training data organized in terms of attributes to be able to represent missing and sparse data occurrences from related sources?

These questions are a follow-on to the question in Section 3. Not only is the correct proportion of training data needed, but the proportion must be in alignment with the reality of the system being deployed. In other words, what the algorithm will experience if involved in a mission.

- How many operational use cases were created and then translated into training data requirements?
- Were the data sources feeding the input to the AI adequately assessed and how does anyone know?
- What intelligence sources were used and how reliable were those sources?

Creating a training mix means that the developer is assuming that the algorithm will need to perform in an imperfect world, and some of the primary sources for the algorithm may not exist. Were secondary sources considered or even tertiary sources? Primary, secondary, and tertiary sources are considered mixes within the training set and can address the missing or sparse data reality during deployment.

- Missing Data: This refers to the data input to the AI model. For our purposes, missing data occurs when the model is expecting certain features but does not receive them because of an issue with the data collection mechanism feeding the model. For example, a sensor states a ship is moving at 1000 knots and therefore has been considered erroneous data. In this case, velocity is considered missing, reported as an empty field in the input stream. The missing data feature comes from some form of data collection failure and can be represented in a field as a blank field, i.e., no data shown. This causes a need for secondary or tertiary attributes mixes.
- Sparse Data: This refers to the data input to the AI model. For our purposes, sparse data occurs when the model is expecting certain features but doesn't receive them, but not because of any issue with the data collection mechanism feeding the model. In other words, sparse data occurs when the system is working but no data is available to fill a field. An example of sparse data might be a fully functional radar system not receiving any blips because there's no target to reflect back. Most times sparse data will be represented by a zero where as a blank field represents missing data. This causes a need for secondary or tertiary attributes mixes.

Part 4—Mission Alignment: How are we ensuring that the algorithm being deployed, after using training data, provides the correct answer when data input issues occur? This question relates to understanding how the training data will be created/curated with regard to potential deployment errors represented by the training data.

- Is analysis of the algorithm's success rate a function of attribute availability within its anticipated operational environment?"
- Will the training set represent the reality of data input issues during deployment? If so, then how will the success rate be affected, i.e., will the success rate be assessed; before errors, without operational issues, after errors are injected, or with operational issues?

Part 5—Oversight: Can other control entities (such as a human operator) be inserted into the loop to reduce the autonomy? One way to answer this question is through Interdependency Analysis (IA). IA allows an objective review to determine which function is best performed by automation or a human operator to create an optimal relationship. The approach helps to optimize performance and understand how best to reduce autonomy with human oversight or guardrail/gate control of critical functions. Requirement content needed for an IA analysis should include:

- Identification of AI enabled functions at the subsystem composition detail.
- Identification of performers, both machine and human, involved with that function.
- Identification of the method(s) used to ensure the interaction between the human and the machine in terms of observability, predictability, and directability.
- Description of the multiple paths through the key function where applicable.
- Description of how necessary metrics can be obtained to objectively support any subjective determinations to reduce autonomy discovered through the IA process.
- IA failure walk through, including any failure modes associated with AI/ML enabled functions.

Part 6—Sparse and Missing Data**:** There are three subparts to sparse and mission data content questions and considerations.

Subpart A: What are the ratio requirements of *sparse* and *missing data* occurrences to normal operations when creating training data from synthetic or live data?

Assuming that sparse and missing data are part of the training data, this question focuses on an expected ratio of occurrence in an operationally deployed environment. If an *Instance* consists of attributes that the AI algorithm is learning to analyze; and *Sparse* and *Missing data* indicate noise in the attributes, making it more difficult for the algorithm to perform, then what ratio of noisy instances make up the training data? This should be a ratio that can be measured for validation and defined in a requirements document. It should not be left to the developer or to chance. Once a ratio is determined, the developer should have confidence, whether it be synthetic data or live data, that it will perform as defined.

Subpart B: Will there be secondary or tertiary attributes supporting the mission or sparse data issues? In other words, if primary attributes are not available for algorithm analysis, will less important attributes be available, e.g., background environment or habit factors? When primary attributes are unavailable due to potential real-world issues, secondary or tertiary sources can increase the success rate of an algorithm's analysis. Therefore, they should be considered when defining a training data ratio.

Something else to consider when determining secondary or tertiary attributes and related types of ratios for:

- Modality 1: How are higher priority attributes that experience sensor malfunction, message corruption and human input errors being mitigated by forcing mixes of lower-level attribute training data to ensure the algorithm deals with "real" operational issues?
- For Modality 2: If there is corruption in parts of an instance, e.g., a blurred image, especially containing higher priority attributes, are secondary and tertiary attribute mixes of training data ensuring the algorithm can deal with "real" operational issues?
- For Modality 3: Are combinations of modalities 1 and 2, regarding training of the algorithm, able to deal with "real" operational issues?

Subpart C: Will the architecture, design and code support sparse and missing data management, or more specifically, will it filter or use a selection of less significant attributes to do the calculations? Note: This question provides discussion regarding the mix of data and how the architecture, design and code will support this mix.

- How will the effects of *missing* and *sparse data* be minimized within the architecture, design and code, from a requirements point of view?
- Will secondary and tertiary attributes be included in the training data, and if so, will secondary and/or tertiary attributes be used as a way to deal with missing and sparse data? If this isn't considered and potentially included as a requirement, it may cause poor success rate performance during deployment of the algorithm.

If this consideration becomes a requirement, implementation of an approach to deal with this issue should be traced throughout the process of development.

- For Modality 1: Will sensor, communication link or human input content elements take priority over the others to improve the success rate when training a ML algorithm under normal to stressed operational conditions?
- For Modality 2: Which attributes, within the single data source, take priority for improving the success rate when training the ML algorithm under normal to stressed operational conditions?
- For Modality 3: What data source content is more significant, with regards to normal to stressed operational conditions? When dealing with separate streams, which of the following: sensor, communication link or human input content elements takes precedent, for improving the success rate when training a ML algorithm under normal to stressed operational conditions? When dealing with combined streams, which attributes within the single data source are identified as primary, secondary and tertiary, regarding importance for ML algorithm to improve success rate, under normal to stressed operational conditions?

Part 7—Data Curation: What processes are being defined, to support data management curation, to ensure that the ML algorithm provides accurate data input?

Data Curation. Is the organization and integration of data collected from various sources. Data curation involves annotation, publication and presentation of the data such that the value of the data is maintained over time, and the data remains available for reuse and preservation. Data curation normally supports a targeted machine learning goal, where the organization is based on the classification or regression needs of the algorithm.

- How does your data curation approach avoid "garbage in, garbage out"?
- What definitions will be used to constitute "garbage in, garbage out"?

These questions ensure that the data curation process for handling data and the creation of training data is understood at the requirements phase. The emphasis will be on ensuring that the data curation process can determine, with some level of measurable certainty, whether accurate data input is being achieved.

For all three modalities: What is the priority list (from highest to lowest) of attributes being used for training? How much more emphasis is placed on the quantity of training data variations with a higher priority than lower?

Part 8—Battle Complexity: How well does the particular ML algorithm support increased battle complexity and how does that affect sparse and missing data issues?

Battle Complexity: A situation described by a series of events, caused by actions between opposing participants, where the outcomes can be significantly affected by factors categorized as: (1) “known-knowns” (facts); (2) “known-unknowns” (assumptions); (3) “unknown-knowns” (absent data); and (4) “unknown-unknowns” (surprises).

- “Known-knowns” (facts) are factors that participants depend on as “fact” to win the engagement; these can include own participant’s technical capabilities, geo-spatial, temporal situational awareness, interoperability, tactical actions and strategy pros/cons.
- “Known-unknowns” (assumptions) are factors that each participant needs to “assume” about variations (of the facts) regarding battle conditions, these can include the third-party involvement, weather forecast, kinetic and non-kinetic effectiveness, opponent’s attack surfaces and related vulnerabilities, heroism and initiative on all sides, opponent’s priorities, and difficulty in overcoming manmade and natural obstructions.
- “Unknown-knowns” (absent-data) are factors that cause a participant to be “absent of data,” sometimes decision critical info; these factors can include human mistakes, sensor failures and communication issues.
- “Unknown-unknowns” (surprises) are factors that will “surprise” participants during the engagement; these include unforeseen technology and anything not anticipated in the previous three categories.

Trust is gained through the LOR descriptions. For example, understanding the modality of the training data (as facts); or as will be described in follow-on LOR descriptions, conducting TSAT and StAR-n analysis to support variations to the input (as assumptions) and providing missing and sparse data class tables (as absent-data). The challenge involves the inability to prepare the AI model to handle unbound data issues (as surprises). Unbound data by its inherent definition means that confidence in the performance/behavior of the AI model cannot be predicted and therefore cannot be trusted.

Given the above definition, consider the following when discussing the topic of trust:

- Can we trust that the training data *factually* represents the real world when deployed, e.g., use of correct attributes/features, noise/background, etc.?
- Can we trust that the *assumptions* regarding input variations from the training data are within expected scope as not to cause an error in the output, e.g., miscategorization?
- Can we trust that the *absence of data* when needed to the AI model has been adequately anticipated and compensated to maintain success rate?
- Even if the previous three answers are all “yes,” the AI model, by definition, is not trained to handle *surprise* inputs, i.e., unanticipated, unbounded data! Historically, we can always expect *surprises* in warfare because intel will never be 100% accurate, i.e., expect the unexpected. Unanticipated/unbounded inputs are known to cause the AI model to have radical, undesirable failures. It is a noted limitation in deep network algorithms, i.e., neural networks with many layers.

Autonomy and AI systems are designed to handle some level of “known-unknowns,” based on the “assumptions” about the variation in the training data input, and are challenged with “unknown-knowns,” relating to missing and sparse, “absent data” issues; but it’s the “unknown-unknowns” that create the more significant concerns regarding “surprises” of unwanted behaviors. In order to represent a realistic operational set of training data, complexity of the deployed environment needs to be considered.

- How will this complexity be considered when synthetically creating or finding data to use for training?

- How will the requirements be defined with regard to complexity?
- Will guard rails/gates be used?

Measurement Report

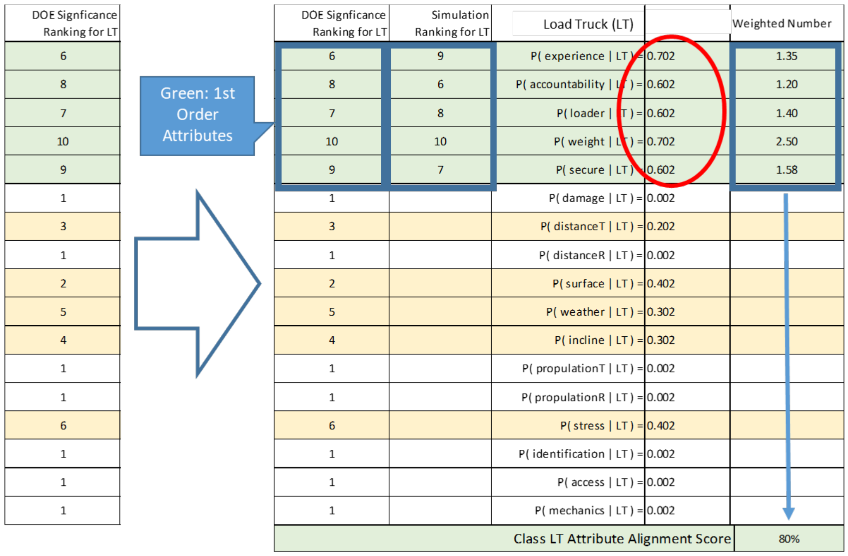
Section 1: Dataset Quality

For each ML class, define requirements that rank the importance of attributes, i.e., creating a priority list, within each instance that the AI algorithm will be trained to recognize. This ranking represents a baseline to determine if a quality training set is being used. As an aid to determining requirements that rank the importance of attributes, a process might be to create operational scenarios looking at nominal and extreme cases. Ranking must be done by class, so the scenarios must be class based with a focus on attribute input to the algorithm.

As an example, a Training Set Alignment Test (TSAT) supports the requirements group in ranking all attributes that will be used by each class. The approach allows the project to assess the training data to determine if the initial ranking is statistically similar to the statistically determined ranking of the training set. Statistical ranking determination is based on a number of occurrences of each attribute within the entire training set. The result of comparing the initial, required ranking to the statistically based ranking is calculated as a single numeric value. The single numeric value represents how well the requirements group's ranking matches the training set compositions. For example, a number of 5.0 out of 10 indicates that the training set only matches 50% of what was required in terms of attribute priority/importance. Having a 100% match is unrealistic, but something above 50% or even 75%, a 7.5 score out of 10, might be a reasonable expectation. The key is ensuring that the attributes within the instances of the training set represent priorities for the algorithm to learn. Priority learning for an algorithm is viewed as how often the attributes and their varying representations repeat. If training on a facial recognition program has only a small percentage of instances that contain nose variations, then the algorithm will not be sufficiently trained to handle and/or properly categorize variations in noses.

- Are attributes for the algorithm ranked in order of priority?
- How does that compare with the actual training data?

These are important questions that need to be addressed and adding these requirements becomes vital to the understanding of whether or not the algorithm is using a quality training set?



DOE Signficance Ranking for LT
6
8
7
10
9
1
3
1
2
5
4
1
1
6
1
1
1

DOE Signficance Ranking for LT	Simulation Ranking for LT	Load Truck (LT)		Weighted Number
6	9	P(experience | LT) =	0.702	1.35
8	6	P(accountability | LT) =	0.602	1.20
7	8	P(loader | LT) =	0.602	1.40
10	10	P(weight | LT) =	0.702	2.50
9	7	P(secure | LT) =	0.602	1.58
1		P(damage | LT) =	0.002	
3		P(distanceT | LT) =	0.202	
1		P(distanceR | LT) =	0.002	
2		P(surface | LT) =	0.402	
5		P(weather | LT) =	0.302	
4		P(incline | LT) =	0.302	
1		P(populationT | LT) =	0.002	
1		P(populationR | LT) =	0.002	
6		P(stress | LT) =	0.402	
1		P(identification | LT) =	0.002	
1		P(access | LT) =	0.002	
1		P(mechanics | LT) =	0.002	
		Class LT Attribute Alignment Score		80%

Figure 5. Training Set Alignment Test (TSAT)

Figure 5 is an example of a TSAT where a Design of Experiments (DOE) ranked a series of 17 attributes supporting 5 classes, LT being a class in the TSAT example, as compared to a series of Monte Carlo simulations that determined the ranking based on the percent/frequency of simulations that used those attributes. The above score for the example is 8.3 out of 10, and would indicate attribute occurrence within the data set are aligned with expected deployed priorities.

Procedure for calculation:

1. Determine a scale for grading from 1 to “m,” where “m” means greatest attribute priority/significance based on operational deployed needs.
2. Identify attributes a_1 to a_n to grade, such that “n” is the number of attributes being graded out of r total attributes available. Therefore n ≤ r and n ≤ m, where grading a_i with grade “m” indicates a_i (m) is the most important attribute based on operational needs. Additionally, attribute grading range is (m-n+1) to m, consecutively, where lowest grade indicates least operationally important (possibly DOE analysis and/or SME determination).
3. Identify the n attributes that occur the most times in the training data. Using the same scale “m,” grade attributes b_1 to b_n based which attribute occurred the most often within the training set (this can be a statistical number, e.g., 70% of the time b_i attribute was used in simulations or 70% of the samples/instances were collected, e.g., images, that contained attribute b_i). Again, grade “m” indicates b_i occurred the most and (m-n+1) indicates b_j occurred the least within the training set.
4. Perform $k = \sum_{i=m-n+1}^{m}(\mathrm{i})$ and $\beta = \sum_{i=1}^{n}\left(\frac{a_i(grade)}{k}\right) * bi(grade) \leq m$
5. Perform $\left(\frac{\beta}{m}\right) * 100 = \alpha\% \geq 50\%$ as a constraint

Section 2: Dataset Quantity

Once attributes are ranked in terms of priority, the next question should be, “Does the ranking indicate a grouping of attributes based on the importance and availability of data during a mission?” In other words, can ranking from *(1 to m)* represent primary attributes or more specifically, are the key attributes that the algorithm depends on used? If so, then attributes ranked *(m + 1) to n* represent secondary attributes. When some of the primary attributes are missing, secondary attributes may be used as input for the algorithm to produce a more successful categorization rate. This also means that a mix of primary and secondary attributes are needed as part of the training data. It should be noted that primary attributes should occur more often than secondary in the training data, based on what is most important for the algorithm to learn.

Primary, secondary, tertiary, and etc., will be based on how often a grouping of attributes are expected as input to the algorithm during deployment. If they were all considered primary, what happens when there is missing and sparse data issues during deployment? Missing and sparse data, by definition, means primary attributes were not available. Therefore, to support realism, should secondary and tertiary attributes be considered? If considered, what should be the ratio of primary to secondary and tertiary attributes? Can this be a requirement?

- Specifically, how will the priority and ratio of a grouping of attributes be determined and how will it be used for testing?

As an aid to determining priority and ration of a grouping of attributes, a process might be to create operational scenarios looking at nominal and extreme cases. Warning, this is only an optional starting point. The focus is on impact to the input attributes to the algorithm, i.e., mission or sparse data events. Manually developing even dozens of scenarios would not be enough. Each scenario manually developed would likely needs fifty or more variations in attribute values for algorithm training. For creating training data, total of all scenarios developed should consist of ten to many hundred thousand variations, balanced by class, as will be discussed. These variations need to be either collected, instances itemized attribute by attribute or instances synthetically created with the goal of creating the desired ratios. For synthetic generation, a Generative Adversarial Network (GAN) inspired approach provides a randomness associated with the creation of large sets of data covering a range of potential issues that the algorithm might encounter during deployment.

The reason why randomness is important is because of the inability to predict future deployed engagement events. Randomness of attribute values within scenarios ensures greater readiness to handle unknown future events. It is because tens of thousands of variations increase the likelihood of the algorithm being trained to handle unanticipated deployed situations. These attribute combinations associated with classes need to be assessed based on their ranking of importance determined earlier.

As an example, the Source to Attribute Ratios for 1, 2, 3 (nth) (StAR-n) Order Matrix approach can support the development of requirements based on attributes (primary, secondary, or tertiary groupings, e.g., a third order matrix) representing highest significance (priority/rating defined in TSAT). Higher priority grouping, e.g., primary, should occur in greater numbers of instances within the training set, by class, than a lower significance grouping of attributes, e.g., secondary. The comparison of numbers can be analyzed as ratios.

Why should developers verify that primary instances have greater numbers than secondary, and so on?

- With live data collection, it might be difficult to find all the training data that includes the appropriate noisy environments that might cause missing and sparse data issues; and
- With synthetic data creation, simulation may be too ideal, not representing the appropriate noisy environments. (Remember that there's most likely an infinite number of possibilities in terms of training data variations and simulation time to meet schedule

demands might get strained.) What should be the priority when considering or designing your synthetically created training data?

The StAR-n Order Matrix focuses on the quantity of how often attributes occur, by their grouping, within the training set. StAR-3 looks at the ratios of primary, secondary and tertiary attributes, as they are defined through requirements.

As stated earlier, training data is key to the AI/ML algorithms development and therefore, the question becomes, “How much of the training data consists of primary vs. secondary vs. tertiary attributes that are dependent on data sources that will be available in the field?” Again, the issue becomes dealing with missing and sparse data during deployed operations.

StAR-n provides confidence that there’s an adequate quantity of training data, whether generated from live events or synthetically created to train the algorithm. StAR-n represents these ratios in the form of a matrix, consisting of three colors, to relate the amount of justification needed to support the training data quantity required. This is similar to a risk matrix coloring scheme. Once the matrix is defined, the actual training data ratios of primary, secondary, tertiary, etc. can be placed within the matrix to determine rigor documentation needed.

The color zones are:

- Zone Green: Evidence of data by showing appropriate n-th order groups of training sets, collected from “live” data or generated by the simulations, including success rates as well as the TSAT results.
- Zone Yellow: Zone Green evidence plus justification of why the n-th group priority can still handle the unexpected and provide acceptable success rates.
- Zone Red: Zone Green and Yellow evidence as to how the algorithm is going to be supervised or monitored when operationally unexpected events occur.

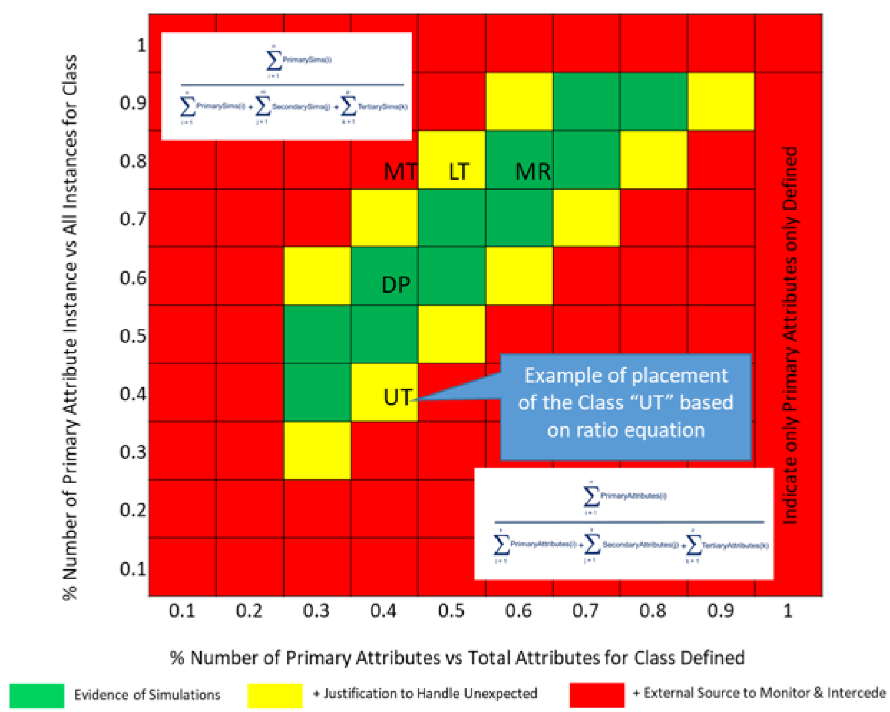


Figure 6. Source to Attribute Ratios for 1, 2, 3, . . . nth (StAR-n) Order Matrix

Above is an example of a StAR-n Order Matrix focused on Primary Attribute ratios. Matrices can be created for primary, secondary, and tertiary attributes, not just primary. Again, zone placement is based on operational needs. Zones can also be changed in terms of how much justification is needed or added ("+"). When measuring actual ratios, placement of the actual ratios would determine what cell the class will be placed, therefore what justification, green, yellow, or red, is needed to support that ratio.

The steps discussed below for taking a StAR-n measurement are groups by requirements/architecture and algorithm code stages:

During the requirements and checked during architecture review:

- First Step: Create a ten by ten matrix, labeling each axis from zero to 1.
- Second Step: Label the horizontal axis "% Number of Primary Attributes vs. Total Attributes for Class" and the vertical axis "% Number of Primary Attribute Instances vs. All Instances for Class."
- Third Step: Determine a three-color zone scheme (as in the example), where green indicates that the ratio fell within acceptable limits, yellow indicates ratio is boarder line acceptable, and red color zone indicated ration is outside expected limits. The color of the zone should indicate how well training data reflects operational environment. Based on color zone, determine evidence justification. Examples (used for guidance only) are described below:
    - Zone Green: Evidence of data by showing appropriate n-th order groups of training sets collected or generated by the simulations, including success rates as well as the TSAT results.
    - Zone Yellow: Zone Green evidence plus justification on why n-th group precedence can still handle the unexpected and provide acceptable success rates.
    - Zone Red: Zone Green and Yellow evidence as to how this algorithm is going to be supervised or monitored when operationally unexpected events occur.

During Algorithm code review when the training set is produced:

- Fourth Step: Calculate the σ and δ (as in the example) ratios. Each ratio should be less than 1. The example below is for primary attributes, but can be done for any n-th order attributes:
    - σ (by Class) = (Number of Primary Attributes / Number of All Attributes) ≤ 1.
    - δ (by Class) = (Number of all Primary Instances / Number of All Instances) ≤ 1
- Fifth Step: Plot (x, y) using (σ, δ) pair of numbers and assess where the pair fall within the color zones to determine support action. An example is provided in the example.
    - Zone Green: Evidence of data by showing appropriate n-th order groups of training sets collected or generated by the simulations, including success rates as well as the TSAT results.
    - Zone Yellow: Zone Green evidence plus justification on why n-th group precedence can still handle the unexpected and provide acceptable success rates.
    - Zone Red: Zone Green and Yellow evidence as to how this algorithm is going to be supervised or monitored when operationally unexpected events occur.

Section 3: Dataset Measurement Review

How do you know if the quality and quantity of Training Data is sufficient? Quality refers to the correct number of attributes (including primary, secondary, etc. mixes) that are representative of the deployed operational environment, including noise factors. Quantity refers to the amount of data/instances used for training, with consideration to mix ratios, underfitting, overfitting and majority/minority classes.

- How do you assess the operational limits described by the training data? (Consider the "You don't know what you don't know" issue.)
- Did the training set include enough noise/clutter for each class (in this case, less significant attributes determined by SMEs for a particular meta-model class) to ensure that the function works properly when deployed? Are there sparse data and/or mission data issues? How is the bias of the training set and variance of the test results determined?

For simulation generation of the training data:

- How would you ensure synthetic or live data configurations work, i.e., is the training data covering the real-world experiences? (Optimizing bias [how well it fits the training set] and variance [how well it predicts using the test set], including considerations of overfitting/under-fitting).
- What quality of synthetic or live training data, i.e., attribute composition on each instance, and how many of these various compositions are really enough to train an algorithm?

Test and Evaluation Readiness Report

Has a process been identified to ensure that randomly selected T&E data is available for testing from the curated training data before any developer uses it? If not, why not?

Will model versioning control be used to track model drift or data drift? Will the versioning control support troubleshooting of any AI model issues that might occur later? Data versioning supports the ability to version multiple sets of data against many different compiled algorithms and then rollback/forward to different training data sets depending on need?

- **Model drift.** Is a form of model decay caused by not keeping the model current with significant attribute changes in the training data, e.g., boys face evolves to a man's face but never updated in the training data.
- **Data drift.** Is undocumented changes to data structures, semantics, and infrastructure, e.g., undocumented modification to the API causing the model to view that part of the input as missing or sparse data.

The model versioning control process should include positive control over who, what, where, and when transactions occur involving the creation of the training data composition.

As an example, the need to use positive control over a training set would be when a T&E set of training data from a k-fold cross validation approach is identified. If live data is limited in terms of quantity available, it is recommended that T&E training data needs take priority and that possibly all live data be set aside just for T&E testing. In either case, there must be a separate amount of training data, randomly selected from a pool of training data that is untouched/unviewed by the developer and specifically focused on supporting T&E.

Since the training data drives the composition of the algorithm during training, it is important that the creation of the data, part of which will become the T&E test set, has strong oversight, in addition to versioning. An approach to provide this oversight, especially when the data is coming from many diverse sources, with multiple touch points, is a technology called blockchain.

Blockchain is a type of distributed e-ledger (similar to what an accountant would use to keep track of financial transactions). It is designed to be a form of tamper-resistant, decentralized documentation that provides proof of transaction involving physical or intellectual assets, in this case T&E training data. It ensures confidence that only people allowed to access the data, from its origin to a T&E facility (separating this test set from the development test set), have access to the data.

By using a blockchain approach, policy enforcement can be ensured and that the rules for accessing the data are followed. A blockchain architecture documents the who, what, where, and when user access (transactions) involving the creation of the data set composition, data attribute transfer to location for T&E random selection, ownership of the T&E test set and integrity of the data.

Missing and Sparse Class Analysis

Will a Missing and Sparse Data (MSD) Class Table, consisting of four sections, be used? The MSD Class Requirements Table provides requirements/guidance for developers to deal with missing and sparse data issues. As part of the requirements, within the table, you can indicate a plus or minus percentage for meeting the numbers listed. As an example, you can have four sections focused on various aspects of missing and sparse data.

If this table is created, an equivalent MSD Class Actuals Table must also be created to be filled in during the development process and then compared to ensure listed requirements are being met.

Using the sandbox use case involving robots delivering package, the tables below provide a five class, 17 attribute example of a package delivery system involving trucks loaded with robots (LT), driving to a drop off location (MT), unloading the robots (UT), having the robots move to the desired location (MT), and deliver packages to the recipients (DP).

Section 1: Class Representations in Dataset

Create a table or list by class for the expected training data quantities/numbers based on ML Training Data. The headings need to describe the "Number of Max Data Sources Allowed for a Decision," "Number of Primary Attributes Based on Data Source Availability," "Number of Secondary Attributes Based on Data Source Availability," "Number of Tertiary Attributes Based on Data Source Availability," and so on. The goal is to have an understanding of data source availability during deployment and the number of attribute inputs (from those data sources) that will feed the algorithm/model.

Table 1. Training Data Attributes Table

Class	Number of Max Data Source for a Decision	Primary Attributes	Secondary Attributes	Tertiary Attributes	+/- % Compared to Actuals
Load Truck (LT)	h1 + h2 + h3	h1	h2	h3	x%
Move Truck (MT)	i1 + i2 + i3	i1	i2	i3	x%
Unload Truck (UT)	j1 + j2 + j3	j1	j2	j3	x%
Move Robot (MR)	k1 + k2 + k3	k1	k2	k3	x%
Deliver Package (DP)	l1 + l2 + l3	l1	l2	l3	x%

Table 1 is an example listing variables h, I, j, k and I that would be converted to numbers supporting requirements for each class. The x% represents the acceptable variance allowed when compared to actual results.

Section 2: Class Ratios of Classes

Create a table or list that describes, within the training set, an expected percentage of how often primary attributes occur in an instance/sample compared to the total number of instances being used for training. Also create percentages for instances of secondary attribute occurrences to the total number of instances, as well as tertiary attributes, etc. These percentages should be defined for each class.

Table 2. Attribute Instances by Significant Grouping Instances Table

Attribute Instances to Total Instances	Classa	Classb	Classc	...	Classn	+/- % Compared to Actuals
Primary to Total	a1%	b1%	c1%		n1%	y%
Secondary to Total	a2%	b2%	c2%		n2%	y%
Tertiary to Total	a3%	b3%	c3%		n3%	y%

Table 2 is an example listing variables a, b and c that would be converted to numbers supporting requirements for each class per priority grouping. The y% represents the variance allowed as acceptable when compared to actual results.

Section 3: Success Rates

Create a table or list that describes the expected success rate when combining attributes from various priority groups of the algorithm (e.g., as a percentage). They can then be measured using the T&E set created from the k-fold cross validation approach described in LOR 8. This description should list the required test results by primary, secondary, and tertiary priority groupings and when mixing groups, e.g., primary only success rate, primary with secondary success rate (with primary as the majority of attributes in the instance), primary with tertiary success rate, secondary with primary (with secondary as the majority attributes in the instance), and so on.

Table 3. Attribute Instances by Significant Grouping Table

Test Results by Class	Primary data source/attributes	Secondary data source/attributes	Tertiary data source/attributes	+/- % Compared to Actuals
Primary data source/attributes	Actuals: 1st Order Only testing success rate (emphasis 1st Order)	Actuals: 1st and 2nd Order testing success rate (emphasis 1st Order)	Actuals: 1st and 3rd Order testing success rate (emphasis 1st Order)	z%
Secondary data source/attributes	Actuals: 1st and 2nd Order testing success rate (emphasis 2nd Order)	Actuals: 2nd Order Only testing success rate (emphasis 2nd Order)	Actuals: 2nd and 3rd Order Only testing success rate (emphasis 2nd Order)	z%
Tertiary data source/attributes	Actuals: 1st and 3rd Order testing success rate (emphasis 3rd Order)	Actuals: 2nd and 3rd Order Only testing success rate (emphasis 3rd Order)	Actuals: 3rd Order Only testing success rate (emphasis 3rd Order)	z%

Table 3 is an example listing variables a, b, and c that would be converted to numbers supporting requirements for each class per priority grouping. The y% represents the variance allowed as acceptable when compared to actual results.

Section 4: Class Balance

Create a table or list that provides an expected majority or minority class analysis of how balanced (equal quantities) the classes are with each other. This is done to avoid data bias

- Data Bias. Occurs when the training data does not equally represent all of the environment where deployed but focuses on a subset. A form of data bias is imbalanced classes. Imbalanced classes means that one class has more training samples/instances and is significantly larger than the others. The class with the larger number of instances is called the majority class and the smaller number of instances is the minority class.

The table or list needs to describe the expected average number of instances, within the training set, for each class. The list should be divided based on the priority grouping of attributes.

As an example of reviewing combinations:

- 1st order only
- 1st and 2nd order (emphasis/more of 1st order)
- 1st and 3rd order (emphasis 1st order)
- 1st and 2nd order (emphasis 2nd order)
- 2nd order only (emphasis 2nd order)
- 2nd and 3rd order only (emphasis 2nd order)

Therefore, focus is on determining what class is a majority or minority class. In most cases, 1st order only, 1st and 2nd order (emphasis 1st rder), may be the only consideration when analyzing each class.

Table 4. Majority and Minority Class Analysis Table

Possible Combinations of Attributes based on Operational Needs	Expected Instances for Training*					Balanced
	Classa	Classb	Classc	...	Classn	
1st Order Only	a1	a2	a3		an	?
1st and 2nd Order (emphasis 1st Order)	a1	a2	a3		an	?
1st and 3rd Order (emphasis 1st Order)	a1	a2	a3		an	?
1st and 2nd Order (emphasis 2nd Order)	a1	a2	a3		an	?
2nd Order Only (emphasis 2nd Order)	a1	a2	a3		an	?
2nd and 3rd Order Only (emphasis 2nd Order)	a1	a2	a3		an	?
1st and 3rd Order (emphasis 3rd Order)	a1	a2	a3		an	?
2nd and 3rd Order Only (emphasis 3rd Order)	a1	a2	a3		an	?
3rd Order Only (emphasis 3rd Order)	a1	a2	a3		an	?
Total Instances by Class:	Σ	Σ	Σ		Σ	Analysis: Balanced or Unbalanced
If Not Balanced, Majority or Minority Class (Based on Availability of Training Data)	?	?	?		?	

Table 4 is an example listing variables "a1" to "*an*" to ensure balanced classes, meaning there are no larger instances, i.e., there are no more minority classes. The desired result would be that the number of instances is basically the same for all classes.

Consideration: A key goal of the last four sections is to ensure that the developer demonstrates a detailed understanding of potential deployment issues that could affect the AI algorithm. This understanding is measured by the composition quality of the training data reflecting operational "realism." When composition quality accurately reflects the deployed operational environment, it results in an improved performance of the AI model under realistic conditions.

The challenge becomes an adversarial network tradeoff. For example, an image recognition system for a smart phone is trained on key facial features. If the owner is wearing a headband, the smart phone may be stumped until the AI is trained to recognize the owner wearing the headband. However, the phones initial inability to recognize unexpected/surprise variations in facial features, e.g., wearing a headband, ensured others were denied unwanted access to phone.

In the four sections previously described, groupings of secondary and tertiary attributes show that the AI model is being trained to handle deployment variations associated with missing and sparse data. These deployment variations are equivalent to training the smart phone to recognize the user when wearing a headband. The concern is whether these types of approaches to increase the quality of data, i.e., using training data to support unexpected/surprise variations in deployment conditions, are also making the AI model more susceptible to adversarial network attacks.

Is the developer considering the balance between versatility, handling variations, and security? If so, there should be formal analysis associated with identifying this balance. The analysis should describe how versatility is generating greater security issues.

If the developer creates the training set as described in each of the four sections, what are the balance considerations between versatility and security? Are they being considered? Balance is obviously an important analysis and emphasizing either "too much or too little" can possibly lead to an issue in the confidence of the behavior of the AI model or the security of the system. This discussion of balance and any related analysis applies to all LOR focused on ensuring quality training data composition

Summary

The development of advanced artificial intelligence (AI)/machine learning (ML) systems for deployment by/throughout the Department of Defense (DoD) is a reality. AI/ML integration into DoD is in the form of upgrades to existing programs or new program acquisitions. How do we know these AI/ML-enabled systems will perform as intended? This paper presented an approach in the form of an AIDP.

Subject "Project Overmatch," Memo October 1, 2020 from the CNO begins, "The Navy's ability to establish and sustain sea control in the future is at risk." In the memo, he goes on to explain that we need to catch up to our competitors in autonomy and AI. The fourth paragraph starts, "Bring me your initial plan within 60 days, and update me every 90 days thereafter." This was a direction to all RDT&E centers to increase acquisition of new AI and autonomous technology. This puts a burden on the acquisition community to ensure reliable AI systems for deployment.

With regard to the DoD's AI deployed systems, there are no policies, guidelines, or tools that ensure reliability is met during the unique AI aspects of software requirements, architecture, design, code, or test. For example, consider that training data creates ML algorithms. There are currently no constraints on how training data is created too learn from "live" operational environments. Current work in this area demonstrates that training data issues can result in significant redo of work, cost overruns, and schedule delays. Addressing this deficiency is necessary. Because of this deficiency gap in AI policies, guidelines, and tools, system safety practitioners as well as the acquisition community members, including SE and PM, have no support/direction in assessing confidence in the behavior of current WD weapons programs that are implementing AI upgrades. Currently, AI is being developed for integration into critical systems within DoD programs of record. Because this deficiency of support/direction is consistent throughout the DoD, many working groups are attempting to understand the unique aspects of AI software requirements, architecture, design, code and test without any suitable safety guidelines. The five types of documentation comprising the AIDP provides some insight into what the developing agency should provide to support reliable, quality development of AI.

1. AI Justification Report—3 Sections: The “AI Justification Report” asks the requestor to establish that the AI requirements being requested are truly a candidate for AI and cannot be met by traditional software or hardware. The process requires the requestor perform tasks in three sections: Section 1, the requestor conducts an AI Type Function analysis to determine if there is an AI algorithm meeting at least one of two criteria. Section 2, the requestor documents the justification for the proposed use of the AI algorithm vs. using a more traditional software, firmware or hardware technique. And, Section 3, document a justification for the AI algorithm’s level of autonomy.
2. Best Practices Report—2 Sections: The “Best Practices Report” focuses on the requestor presenting development questions and how the AI algorithm will perform its function(s). This process consists of two sections. Section 1 asks the ML Training Data modality type represented in the deployed system and data generation process. Section 2 places emphasis on Dataset Structures by asking questions in eight parts regarding: Representation (training data needed), Fitness of Data (overfit and underfit considerations), Mission Alignment (is training data mission aligned), Mission Alignment (is algorithm adequately trained), Oversight (guardrails or gate control), Sparse and Missing Data (expected ratio of occurrence), Data Curation (process defined), and Battle Complexity (can algorithm support increased complexity).
3. Measurement Report—3 Sections: The “Measurements Report” answers the question about quality and quantity of data in an objective format. It presents two examples of measuring the quality and quantity of a dataset to support a reference for the type of adequate response. Section 1 provides a measurement approach to examine dataset quality. Section 2 provides a measurement approach to examine dataset quantity. The third section focuses on how the previous quality and quantity measurements are representative of the deployed operational environment, including noise factor.
4. Test and Evaluation Readiness Report: The “Testing Readiness Report” ensures that during the T&E process, separate datasets, not used or seen by the developer, are available for the test engineer to use. It also briefly discusses terms like model drift and data drift associated with ensuring that the dataset used is up to date. In addition, it provides guidance on configuration management with an example of using blockchain.
5. Missing and Sparse Class Tables—4 Sections: The “Missing and Sparse Class Tables” consist of four tables/sections, where a table’s goals can be compared to tables with actual results. Mixes of attributes when training each class becomes important because of mission and sparse data issues. Section 1 focuses on how the quantity of classes are represented in the dataset. Section 2 looks at ratios, comparing the quantity of classes to other classes. Section 3 provides a table to capture success rate, again a goals and then actuals. And finally, Section 4 compares the quantities of each class to the other to ensure equal representation for algorithm training.

References

Acquisition Notes (AcqNotes). (2022). *Program management tool for aerospace.* https://acqnotes.com/acqnote/careerfields/contract-data-requirements-list-cdrl

Borysowich, C. (2009). Sample contract data requirements list (CDRL). *ToolBox Tech*. https://www.toolbox.com/tech/enterprise-software/blogs/sample-contract-data-requirements-list-cdrl-

Department of the Army. (2021). *Artificial intelligence requirement guidelines and precepts* [White Paper]. Artificial Intelligence (AI) Safety Working Group and the Office of the Director of Army Safety.

DI-MISC-80508B. (2006). *Data item description: Technical report-study/services.*

DOD-STD-1703. (1987). *Military standard: Software product standards.*

DOD-STD-2167A. (1988). *Military standard: Defense system of software development.*

DOD-STD-7935A. (1988). *Military standard: DoD automation information systems (AIS) documentation standards.*

Kendal, A., Das, A. Nagy, B., & Ghosh, A. (2021). Blockchain data management benefits by increasing confidence in datasets supporting artificial intelligence (AI) and analytical tools using supply chain examples. *Proceedings of the Eighteenth Annual Acquisition Research Symposium*, Naval Postgraduate School.

MIL-STD-498. (1994). *Military standard: Software development and documentation.*

Miller, S., & Nagy, B. (2021). Interdependence analysis for artificial intelligence system safety. *Proceedings of the Eighteenth Annual Acquisition Research Symposium*, Naval Postgraduate School.

Nagy, B. (2021a). *Applying generative adversarial network constructs to mission-based simulations to produce "realistic" synthetic training data for machine learning algorithms, GANs overview* [Presentation slides]. Naval Applications for Machine Learning Symposium. NABN581, Naval Information Warfare Center Pacific.

Nagy, B. (2021b). Increasing confidence in machine learned (ML) functional behavior during artificial intelligence (AI) development using training data set measurements. *Proceedings of the Eighteenth Annual Acquisition Research Symposium, Naval Postgraduate School*.

Nagy, B. (2021c). *Tips for applying artificial intelligence to battle complexity, naval applications for machine learning symposium* [Presentation slides]. Fall Naval Applications for Machine Learning Symposium. Naval Information Warfare Center Pacific.

Nagy, B. (2022a). *Fourteen tips to increase confidence in the performance of artificial intelligence (AI)/machine learning (ML) functions during the five stages of development* [Presentation slides]. 2022 National Fire Control Symposium.

Nagy, B. (2022b). *Level of rigor—How to develop quality AI products* [Presentation]. Naval Applications for Machine Learning, NABN708, Naval Information Warfare Center Pacific.

Space and Naval Warfare Command (SPAWAR). (2005). *Software development plan template* (Code 20203 TM-SPP-02 V2.0). Systems Engineering Process Office. https://www.acqnotes.com/Attachments/Software%20Development%20Plan%20Template%20-%20SPAWAR.pdf

Thiebes, S., Lins, S. & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electron Markets, 31*, 447–464. https://doi.org/10.1007/s12525-020-00441-4

Recommending Recommendations to Support the Defense Acquisition Workforce

Dr. Carlo Lipizzi—is a professor at the Stevens Institute of Technology teaching, researching, and consulting on Machine Learning and Natural Language Processing. He has a PhD in System Engineering from Stevens Institute of Technology, an Executive Management Degree from IMD, Lousanne, CH, and a master's in mathematics from the University of Rome, Italy. As director for the Center for Complex Systems and Enterprises, his research focus is on Machine Learning, Data Science and Natural Language Processing. [clipizzi@stevens.edu]

Hojat Behrooz—works as a data scientist with Stevens Institute of Technology, School of Systems and Enterprises. He has joined Stevens with years of knowledge and experience in intelligent transportation systems. Over the years, he has led several major transportation and traffic operation projects improving the safety, environment, and equity in Tehran's megacity. His efforts to design and implement the first BRT lane was recognized and awarded the Sustainable Transport Award at the TRB conference in Washington, D.C., in 2010. Behrooz has a BSc in Computer Engineering and a PMP certificate from the Project Management Institute. He received his MSc in Engineering Management from Stevens in 2021. His MSc dissertation was on machine learning application in surface transportation systems. [hbehrooz@stevens.edu]

Michael Dressman—is currently a second semester graduate student in the Master of Engineering in Systems Analytics program at Stevens Institute of Technology. He received a Bachelor of Science in Computer Science at Lehigh University with minors in Data Science and Economics. Currently a research assistant in the School of Systems & Enterprises, Michael is interested in data analysis and modeling along with natural language processing and understanding. [mdressma@stevens.edu]

Arya Guddemane Vishwakumar—is currently a second semester graduate student in the Master of Science in Computer Science program at Stevens Institute of Technology. He received a Bachelor of Engineering and Computer Science from the Dayananda Sagar College of Engineering, Bangalore. He has experience in Machine Learning and in front-end development. Currently a research assistant in the School of Systems & Enterprises, developing the user interface for this project.

Kunal Batra—is Manager, Information Systems and Technology for the Systems Engineering Research Center as well as for the School of Systems and Enterprises within Stevens Institute of Technology. He is responsible for the overall planning, organization, and execution of all information technology within the school. He analyzes and manages servers, systems, and assets and recommends and implements solutions. He also carries out support and maintenance of existing applications and development of new technical solutions, leading the developers on software projects serving as a liaison between business and technical aspects of each stage. [kbatra@stevens.edu]

Abstract

This paper presentings the preliminary results of a research study to support the Defense Acquisition Workforce with a Natural Language Processing (NLP)/Machine Learning (ML) prototype of a system to determine what are the most relevant recommendations that stakeholders are providing to the Defense Acquisition community.

The problem addressed by the research study is in the realm of NLP and ML and it is part of the quite popular category of "recommendation systems." Unlike the majority of the cases in this category, though, this task does not focus on numerical data representing behaviors (like in shopping recommendations), but on extracting user-specific relevance from text and "recommending" a document or part of it.

In order to identify important pieces of these texts, subjective text analysis is required to be run. The method used for the analysis is the "room theory framework" by Lipizzi et al. (2021) which applies the Framework Theory by Marvin Minsky (1974) through the use of text vectorization. This framework has three main components: a vectorized corpus representing the knowledge base of the specific domain (the "room"), a set of keywords or phrases defining the specific points of interest for the recommendation (the “benchmarks”) and the documents to be analyzed. The documents are then vectorized using the “room” and compared to the “benchmarks.” The sentences/paragraphs within a given document that are most similar to the benchmarks, and thus presumably the most important parts of the document, are highlighted. This enables the DAU reviewers to submit a document, run the program, and be able to clearly see what recommendations will be the most useful.

Keywords: Recommendation Systems, Contextual Understanding, Text Mining, Natural Language Processing, Text Vectorization

Introduction

Analyzing text and extracting actionable elements from it is intrinsically challenging, as this task is strongly supported by human common knowledge, and therefore automatic systems fail in true semantic understanding.

Traditional approaches to text analysis are based on “Symbolic” processing, where predefined structures (ontologies and taxonomies) are used to extract semantic elements. The problem with those systems is in the limited context-dependent analysis they can perform, being based on structures that are rarely optimized for the specific need and the given time. This is a “rationalist,” rule-driven approach.

Emerging and new approaches are based on heavy use of Machine Learning and employ complex “deep learning” systems inspired by the human brain structure. The problem wtih those systems is not taking into consideration how humans represent their knowledge and how we achieve the understanding of a problem. This is an “empiricist,” data-driven approach.

Our approach is a combination of Symbolic and Machine Learning, with an additional layer of user interface and visualization, to make the findings more usable by the Defense Acquisition Workforce.

The prototype is based on previous projects we developed for the DoD over the last few years, employing a team of 25 researchers and relying on theories and components we developed for those projects.

For the development of the prototype, we focused on 1) creating a symbolic model for the text understanding and 2) design the process to apply it.

The symbolic model is the “room theory framework” by Lipizzi et al. (2021) which applies the Framework Theory by Marvin Minsky (1974) through the use of text vectorization. This framework has three main components:

- a vectorized corpus representing the knowledge base of the specific domain (the “room”);
- a set of keywords or phrases defining the specific points of interest for the recommendation (the “benchmarks”);
- the documents to be analyzed.

The documents are vectorized using the vectors in the “room” and compared to the “benchmarks.” The sentences/paragraphs within a given document that are most similar to the benchmarks, and thus presumably the most important parts of the document, are highlighted.

The process is a set of logical steps including:

- document ingestion from pdf to text via either the graphical user interface or from existing files;
- text cleaning and "n-gramming" (extracting logical elements composed by multiple words);

- rearranging the document in logical paragraphs;
- vectorize the corpus/knowledge base;
- compare documents to be analyzed with the benchmark elements,
- highlight the most relevant sentences/paragraphs in the original documents;
- present the results via graphical user interface.

Literature Review

Recommendation systems are known commonly to be used to recommend what product you should buy or what movie/show/video you should watch. These systems typically use market basket analysis also known as association rules (Agrawal et al., 1993). Having history on what a specific person or account has consumed helps the systems guess what they would consume next. Items that are consumed together or within a short time frame apart are often considered similar conceptually. If one buys peanut butter, they will likely buy jelly. If someone watches *Star Wars: A New Hope*, then recommending that they watch the sequel *Star Wars: Empire Strikes Back* will more often yield in positive results. Along with products and media, text can also be recommended. While existing techniques for processing text are based on retrieving facts, processing of subjective information is still developing. For subjective analysis Machine Learning is commonly used. Opinion Detection (Jimenez-Marques et al., 2019), sentiment analysis (Pinto & Maurari, 2019), and the use of fuzzy rules to improve text summarization (Guolarte et al., 2019) are all examples. Li et al. (2019) used subjective queries for databases, while Wu et al. (2019) developed an algorithm to account for subjectivity in crowdsourced label aggregation.

Finally, a study from 2006 (Lin et al., 2006) highlighted the need for a perspective analysis when detecting subjectivity in text. This line of study became known as stance detection and is commonly used in opinion mining, to identify if the author is in favor or against the object being analyzed (D'Andrea et al., 2019).

Similarity plays a big role in textual recommendations. With extractive text summarization, text is compared to itself in order to find the sentences or paragraphs that are the most similar to all the other sentences/paragraphs. This technique allows text to be represented by a subset of itself without losing too much meaning. Unfortunately, this technique does not work well with large amounts of text which rules it out of the possibility of being used to recommend parts of lengthy DAU documents.

Existing techniques on textual information processing concentrate on mining and retrieval of factual information (e.g., information retrieval, text classification, text clustering, among others). On the other hand, the processing of subjective perceptions, such as emotions, opinions and summarization, is still a developing field. In particular, because of the intrinsic subjectivity of the summarization process, a generalized summarization model has never been developed.

The automatization of subjective/context dependent tasks is not new in Natural Language Processing. Many efficient algorithms, tools, and techniques have been developed in the past few years and can deliver reasonable results. More recent studies appear to focus on improving these existing methods or creating frameworks that combine them for a certain application.

No one of the above methods, techniques, algorithms could be fully applied to our task. We then opted for an approach—the "room theory framework" by Lipizzi et al. (2021)—which provides a framework to be used to address the needs in our task.

The Room Theory developed by Carlo Lipizzi (2021) is based on Mavin Minsky's Framework Theory (1974). In Minsky's theory, he said that a frame is like a data structure that can express/simplify a concept of being in a room. Lipizzi's theory adds onto this the idea of having a computational version of semantic rooms for Natural Language. A "room" represents the knowledge of a specific domain, it has been created from large corpora related to that domain and transformed into vectors for the analysis.

The main idea is to be able to identify certain structures that would classify the document as belong to a specific domain. In this particular case it is to find recommendations inside recommendation documents. The theory leverages "benchmarks" that are keywords or phrases and finds similarity withing the documents to those benchmarks. The benchmarks are curated by subject matter experts to be able to identify relevant sentences/paragraphs. The overall process of the room theory is displayed in Figure 1 below. Documents are used to train a vectorization model to represent each document as an array of word vectors. Documents can then be checked for similarity with the benchmarks and be classified as important/relevant based on the similarity scores. For this task, we used Word2Vec by Mikolov et al. (2013) as vectorization model.

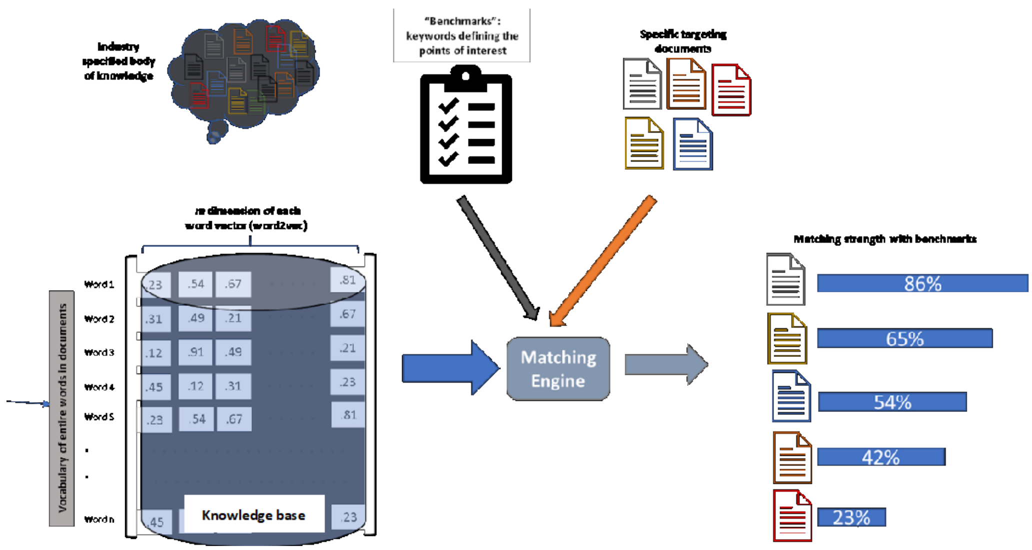


Figure 1. Room Theory Process

Method

To create the room representing the knowledge base for this task, we used a corpus we collected for a previous DAU project. The corpus is composed by1493 pdf files with a total size of 3.1 GB. They were used as domain-specific materials which contained documents related to the DoD and DAU in general. The text was retrieved from these documents using the python library Fitz (PyMuPDF, 2022). Once the pdf files were read in text format, they were then passed to the preprocessing phase. In this phase, the documents were tokenized into words and then preprocessed with steps that include the removal of extra spacing, punctuation, digits, and non-English characters as well as the creation of bigrams and trigrams to be considered as single logical words. After the preprocessing phase, the cleaned word list of each document has been forwarded to a vectorization modeler (Gensim's word2vec) with the following hyperparameters.

w2vec_model = Word2Vec(min_count=10,

```
                window=7,
                vector_size=300,
                sample=6e-5,
                alpha=0.03,
                min_alpha=0.0007,
                negative=20,
                workers=cores-1)
w2vec_model.train(docs, total_examples=w2vec_model.corpus_count,
            epochs=50, report_delay=1.0)
```

The model's output contains word embeddings (word vectors) for 27,229 unique words and n-grams after being trained on 7,826,687 total words and n-grams from the entire input documents. The vector size was 300 dimensions. A sample vocabulary word with their frequency is presented in Figure 2 as a word cloud.

Figure 2. Wordcloud For the Trained Model Vocabulary

In order to create the "benchmark," we assembled a list of words defining the elements of interest for recommendation. The list has been developed with subject matter experts. The initial list is filtered and enhanced with synonyms and misspellings, and then a weight value between 1 to 5 is added to the list. This weight shows the importance of the benchmarks for the targeted subject. Each "word" in the benchmark is actually a list of words, with the original word as a root and additional words being synonyms and misspellings, to improve the benchmarking process. For example, for the word *optimization* would have also terms such as optimizing, optimization, optimizations, optimums, optimizes, optimum, optimizations, optimized, optimize, optimally, optimize, and optimal. By applying this technique to 173 initial benchmarks, 1196 total benchmarks and their roots are created as the primary benchmark list; 423 items of this benchmark are not defied in the trained vocabulary and ignored for further processing. Therefore, the final benchmark list consists of 773 known benchmarks by the vocabulary.

Table 1. A Sample of the Benchmarks and Their Weights

Seq	Benchmark	weight	Seq	Benchmark	weight	Seq	Benchmark	weight
1	accurate	1	31	feasibility	2	61	priority	1
2	achievable	1	32	finalize	2	62	programs	1
3	achieve	1	33	framework	1	63	progress	1
4	act	2	34	goals	1	64	quality	1
5	address	2	35	guidance	2	65	rebuild	2
6	advantages	1	36	identified	1	66	recommend	5
7	align	2	37	implement	2	67	recommendation	5
8	analytical	1	38	improve	1	68	replacement	1
9	assess	2	39	incomplete	1	69	require	1
10	assessment	2	40	incorporate	2	70	requirements	1
11	assignment	1	41	integrated	1	71	revise	4
12	baseline	1	42	lack	2	72	revised	4
13	capability	1	43	lifecycle	1	73	risks	1
14	capture	1	44	maintain	1	74	root cause	1
15	challenges	1	45	manage	1	75	schedule	1
16	challenging	1	46	measure	1	76	setting	1
17	completed	1	47	metrics	1	77	should complete	3
18	conduct	1	48	missions	1	78	should follow	3
19	configured	1	49	modernize	2	79	strategy	1
20	construct	1	50	monitor	1	80	structure	1
21	coordinate	1	51	monitoring	1	81	sufficient	1
22	costs	1	52	moving forward	1	82	support	1
23	critical	1	53	needed	1	83	sustainment	1
24	define	1	54	operational	1	84	system	1
25	develop	1	55	optimization	1	85	take action	3
26	development	1	56	performance	1	86	transition	1
27	effort	1	57	plans	1	87	update	1
28	emphasizes	1	58	policy	1	88	weaknesses	1
29	evaluation	1	59	practices	1			
30	execute	1	60	prioritize	2			

Together with the benchmark list and its weights, the trained model provides the essential tools to evaluate the input documents.

The model can analyze a new input document and measure its specificity from several angles. Two more relevant evaluation process is presented here, which are more applicable to the recommendation system.

Relevant/Irrelevant Input Document

This concept gives a measure related to each input document which shows how much the input text file is relevant to the benchmark. In other words, this measure shows the similarity of the content of the input document to the benchmarks in total. This measure would be a value between 0 and 1, in which higher values show more similarity. A threshold should be assumed to separate relevant/irrelevant input documents regarding the benchmarks. In this process, we used the cosine similarity to create a matrix of similarity between the benchmarks and entire input document words, as shown in Figure 3. Each matrix column represents the whole input document's similarity to a specific benchmark. The frequency distribution of this vector shows how input document words are related to that specific benchmark. A skewed distribution shows that more words in input documents have similarities to the particular benchmark and vice versa.

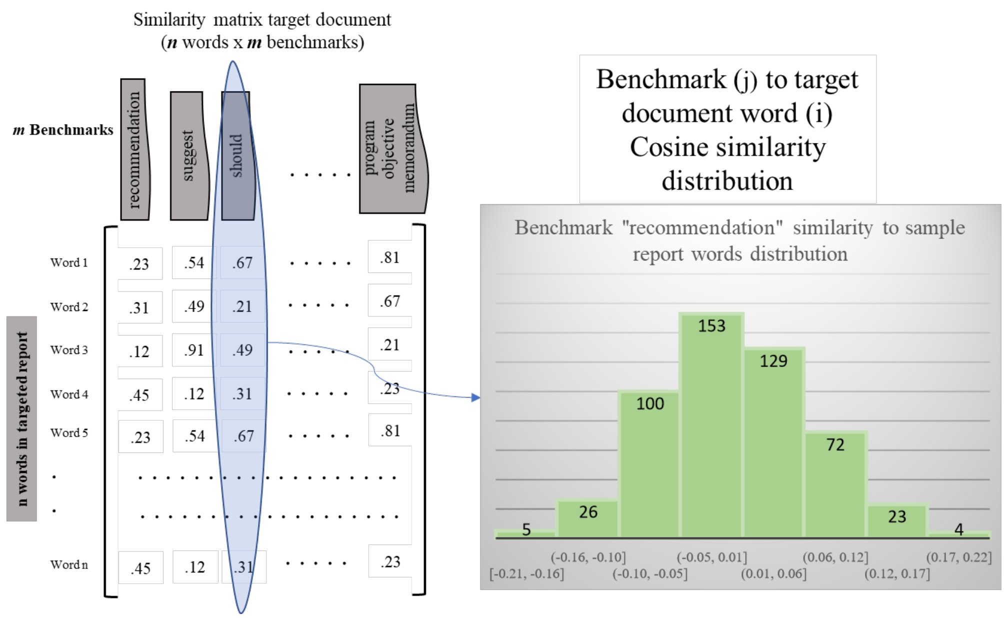


Figure 3. Frequency Distribution of Document's Words and Specific Benchmark's Similarity

A two-bin distribution could be applied((-1,0) and (0,1)). The first bin shows how dissimilar a specific benchmark is to the entire input document. Bin 2 also shows how similar the input document is to the specific benchmark. This second bin could be used to measure the similarity between the benchmark and the entire input document. All the values from the comparison of benchmarks and the input document words create a similarity vector between the benchmarks and the input document, as represented in Figure 4. This will be used to highlight the document/parts of document that are more relevant.

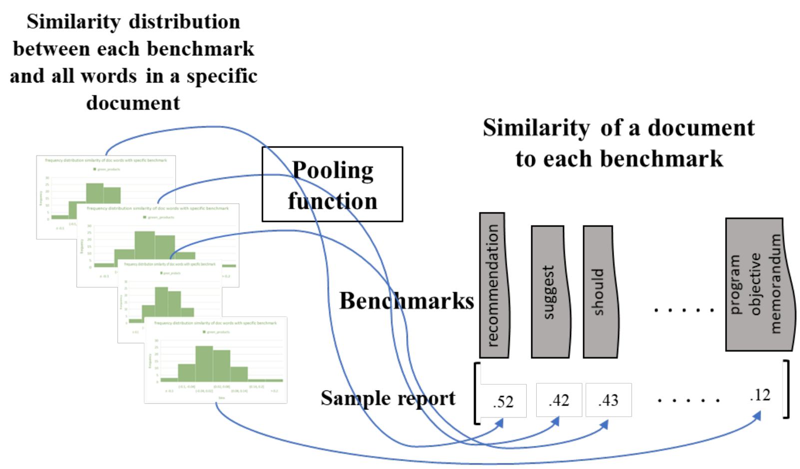


Figure 4. Bin 2 Pooling to Calculate Each Benchmark and Entire Document Words

Another element that can be extracted is drilling down the relevance of each word in the benchmark for each document or part of it. A sample result of the small set of benchmarks and input documents is presented in Figure 5.

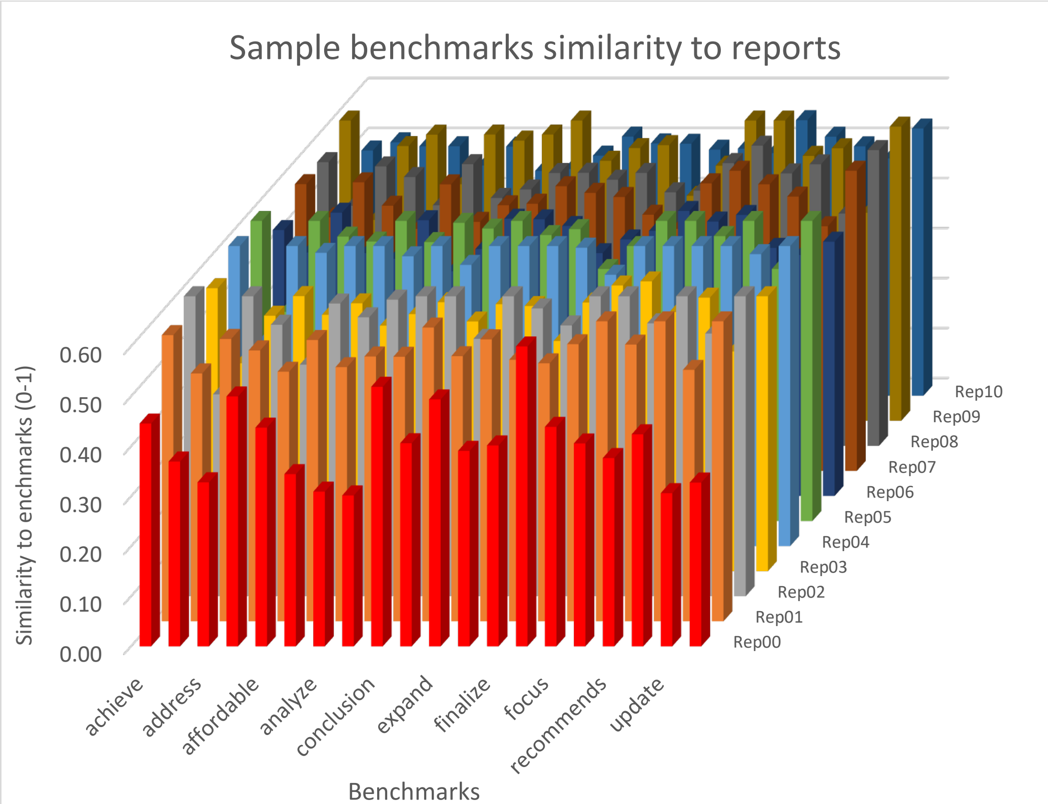


Figure 5. The Similarity of Entire Words of Documents and Each Benchmark's Comparison

As it can be seen, the Rep00, in general, is less similar to the sample benchmarks, while the Rep04 has more similar measures to the benchmarks. By counting the number of words with a similarity more than a threshold (such as 0.50) and normalizing it with respect to the number of the words in each document, an average similarity measure is calculated that presents the level of similarity of the entire document with respect to the entire benchmarks. This single measure for each document can be used to compare various documents to each other in similarity to entire benchmarks. A document with a higher measure is more relevant or like the benchmarks. Figure 6 presents a comparison presentation together with an assumed threshold (0.60) for finding relevant/irrelevant input documents. This would provide an overall view of the documents in terms of their relevance.

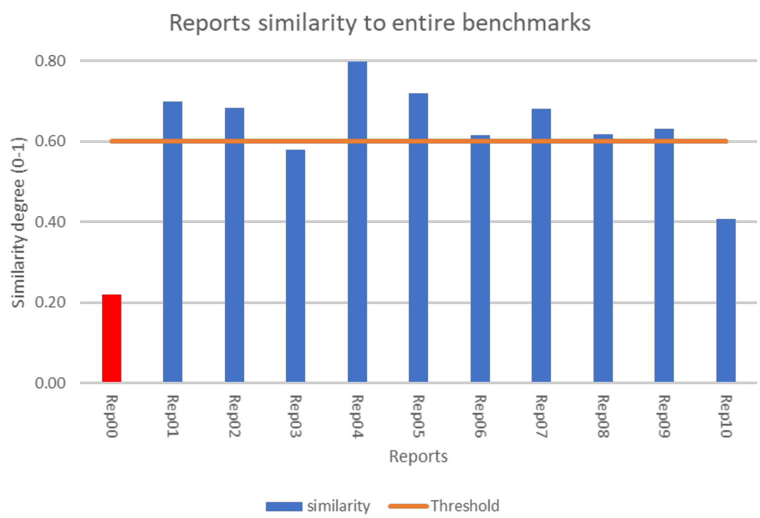


Figure 6. Comparing Various Documents' Total Similarity to Entire Benchmarks

Highlighting the Recommendation Part of an Input Document

As mentioned, we created a similarity matrix for each word in the document and benchmark. The similarity between word n and benchmark m will be from -1 (most dissimilar) to 1 (most similar). The max-pooling technique is then implemented so that there will be an array of length n that will have the maximum similarity between a specific word with the set of benchmarks. Max-pooling is sample-based discretization process to down-sample an input representation by reducing its dimensionality. This is portrayed below in Figure 7.

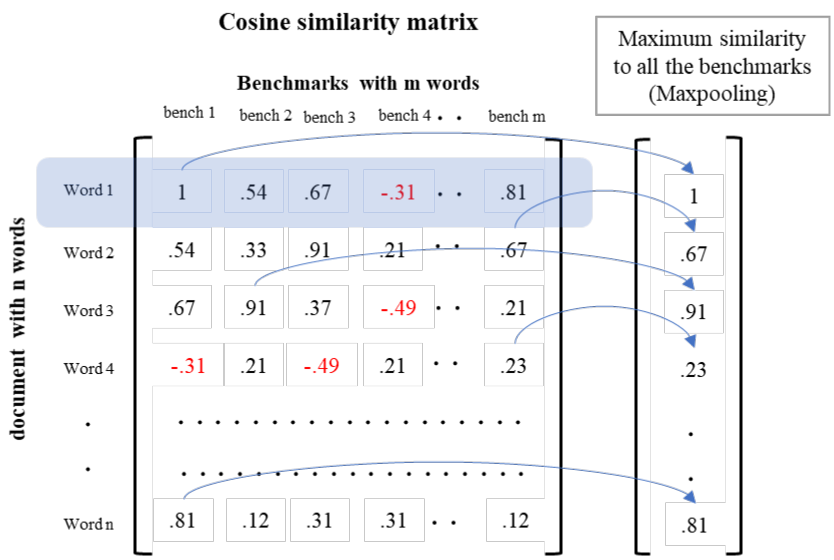


Figure 7. Cosine Similarity Between Document's Word and Various Benchmarks

This max pooling is then weighted, so the max similarity of the word and benchmark is multiplied by the benchmark's weight. So, for example, for word 1, since the maximum similarity is with benchmark 1, the first element of the array would be multiplied by benchmark 1's weight.

To determine the relevant parts of each recommendation, the document was looked at in segments of words. Since words form sentences and sentences form paragraphs when you have relevant words that could be useful for recommendations, the sentences and paragraphs of those words become essential. To account for this, each document's max similarity array had its moving average with a window size of 20 words calculated at each word. In Figure 8, the location of the window of 20 words that maximizes the moving average is shown on a plot of the weighted max pooled similarity over the time of the whole document.

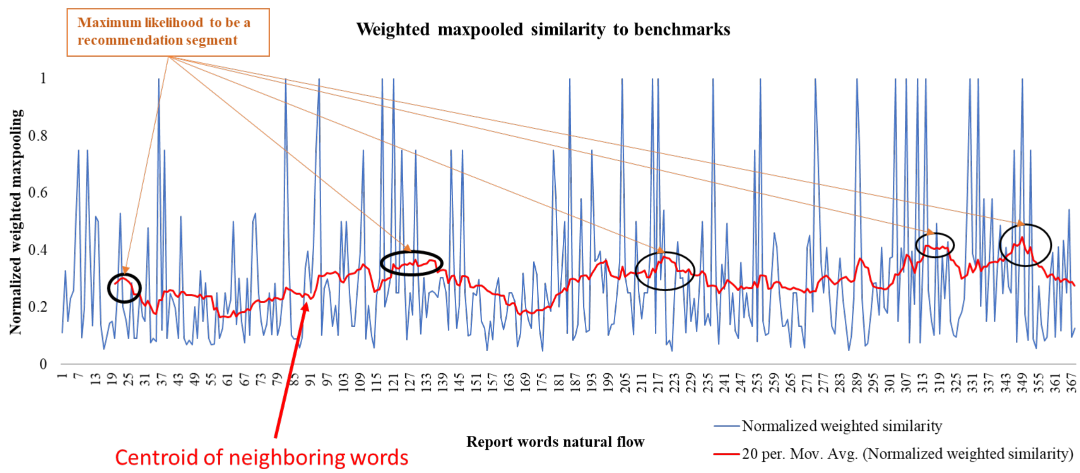
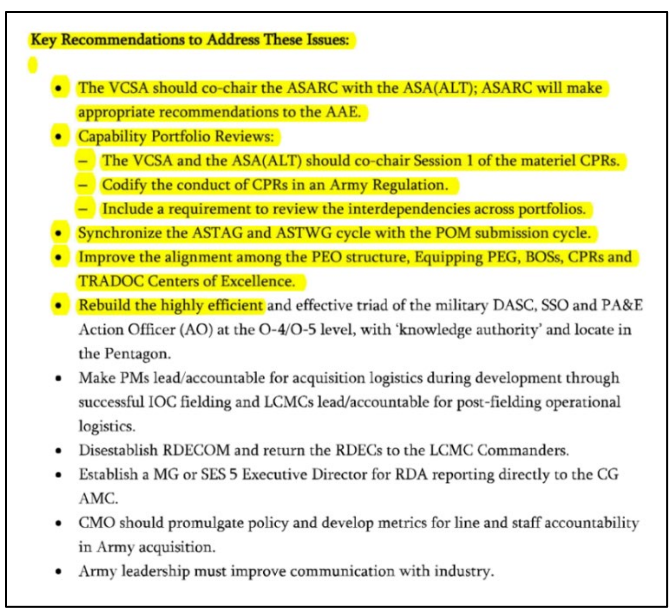


Figure 8. Document Moving Average Similarity to Entire Benchmarks

It was found that with a window of 20 words from the similarity matrix, the actual document (which includes the raw text) would have a window of 35 words that would make up important and relevant recommendations. To assure high-quality moving average windows, the threshold of average similarity is set to 0.75. Any window of words above that threshold is then traced back to the original document and is highlighted. We used PyMuPDF from the Fritz library for this task. An example of what the highlighted section looks like is in Figure 9.


Key Recommendations to Address These Issues:

- The VCSA should co-chair the ASARC with the ASA(ALT); ASARC will make appropriate recommendations to the AAE.
- Capability Portfolio Reviews:
  - The VCSA and the ASA(ALT) should co-chair Session 1 of the materiel CPRs.
  - Codify the conduct of CPRs in an Army Regulation.
  - Include a requirement to review the interdependencies across portfolios.
- Synchronize the ASTAG and ASTWG cycle with the POM submission cycle.
- Improve the alignment among the PEO structure, Equipping PEG, BOSs, CPRs and TRADOC Centers of Excellence.
- Rebuild the highly efficient and effective triad of the military DASC, SSO and PA&E Action Officer (AO) at the O-4/O-5 level, with 'knowledge authority' and locate in the Pentagon.
- Make PMs lead/accountable for acquisition logistics during development through successful IOC fielding and LCMCs lead/accountable for post-fielding operational logistics.
- Disestablish RDECOM and return the RDECs to the LCMC Commanders.
- Establish a MG or SES 5 Executive Director for RDA reporting directly to the CG AMC.
- CMO should promulgate policy and develop metrics for line and staff accountability in Army acquisition.
- Army leadership must improve communication with industry.

Figure 9: Highlighted Text in the Original Input File as a Relevant Part

User Interface

In order to make the system easy to be used by the Defense Acquisition Workforce, we developed a web-based graphical user interface. The interface gets the data from a repository where the documents would be placed and were the temporary results will be hosted. We use a MongoDB database for the repository. The webpage with the user interface contains two separate groups, one for a user and one for an admin. The user logs in with an email and password and they will land at a drop box page shown in Figure 10.

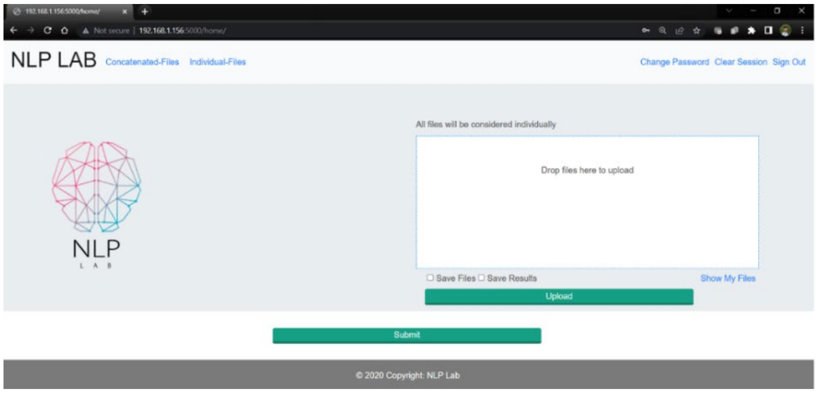


Figure 10. User Homepage

In this drop-box, the user can upload files to the repository for storage. Once uploaded, the file can also be submitted, which would trigger the file being used to run in the recommendation system to get the document similarity to the benchmarks as well as having the document's sections with a moving average similarity over the threshold being highlighted and shown to the user. The user has the option to submit individual files for running through the system or to submit several files that will be concatenated run through the system at the same time.

The admin has the ability to create new users and to manage the users files. Which entails managing the repository on the MongoDB database. In the database, the files are stored in small chunks, as represented in Figure 11.

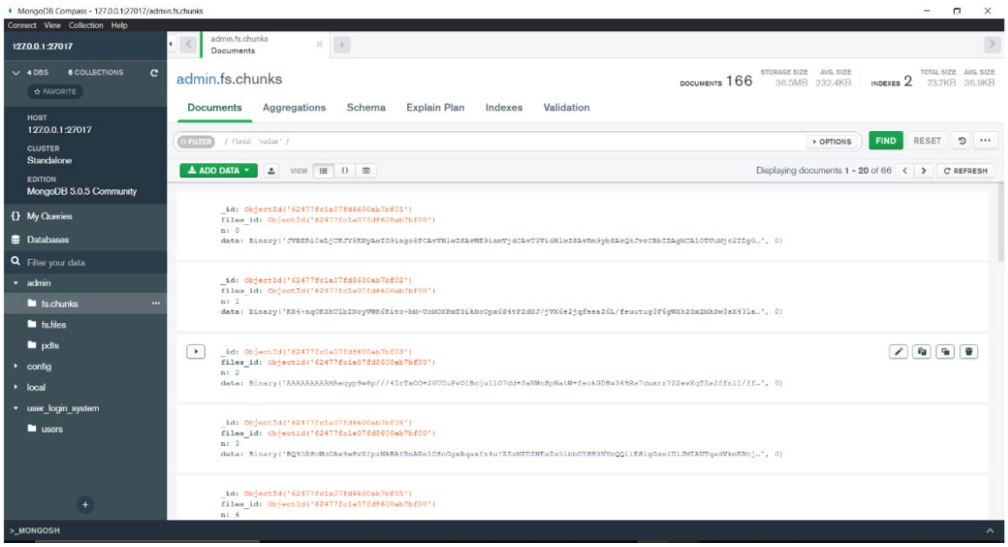


Figure 11. MongoDB Database for File Chunks.

Results

From using 1493 pdf documents and 773 benchmark keywords and phrases in training, 11 documents of varying length were used to evaluate the model. Ten of these documents are recommendation documents, while there is one control document, CMH_Pub_72-2.pdf. Overall, the recommendation documents came back with high similarity regarding the domain-specific benchmarks. A good indicator that the model learned is that the control document's similarity (0.25) was significantly lower than the worst recommendation document (0.5). This means that the model did an accurate job of learning the domain of recommendations and finding the parallels in the documents.

Table 2. Document Similarities Measure to Benchmarks

Input Document	Similarity Degree
AIA - Acquisition Rebalancing 2May2014.pdf	0.80
APMP - Closing the Proc Gap Survey_v6 2014.pdf	0.78
Birkler et al. 2010 - Marginal Adjustments to Meaningful Change - Rethinking Acq RAND_MG1020.pdf	0.66
CMH_Pub_72-2.pdf	0.25
Decker-Wagner Army Acquisition Review - Summary and Implementation 2010.pdf	0.82
GAO - Defense Acquisitions - Where Should Reform Aim Next 29Oct2013.pdf	0.81
Goldwater Nichols - Perfect Storm - Nemfakos Blickstein 2010 RAND OP-308.pdf	0.70
NDIA - Pathway to Transformation Acquisition Report 14Nov2014.pdf	0.78
PSC - Acquisition and Technology Policy Agenda - 28July2014.pdf	0.76
Schmidt 2000 - Acq Reform in US Army - Changing Bureaucratic Behavior - RAND MR-1094-A.pdf	0.70
Sec809Panel_Vol2-Report_June18.pdf	0.50

When the documents' similarity is plotted against the page length in Figure 5, there appears to be a negative correlation. The longer the document is, the lower the similarity score. One exception is the second-longest document, Decker-Wagner, which was 246 pages long and scored the highest similarity. Since there are only 10 data points, it is hard to generalize this to every document, so to understand better if this trend is common, more documents would need to be evaluated by the model. We are also implementing a paragraph-level analysis, as detailed in the conclusions/future development paragraphs. This would provide a better level of granularity in the recommendation: in a longer document there may be parts that are highly relevant, along with others—eventually many others—that are not. This would make the whole document relatively low in relevance, losing the relevance of its key parts.

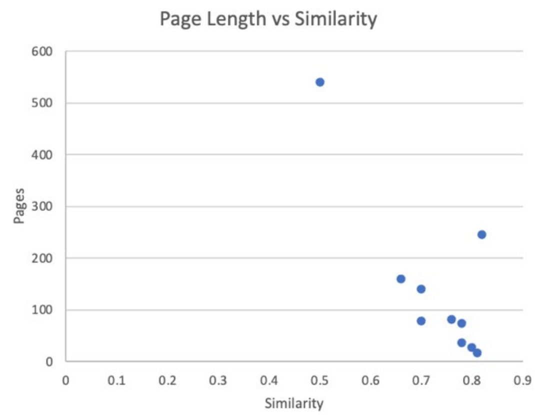


Figure 12. Page Length vs. Similarity

For individual documents, the sentences/paragraphs with high moving average similarities can be highlighted to give the review an easy way to locate the relevant and vital parts of the text. Looking back to Figure 9, the model does a good job of highlighting specific recommendations to be implemented, but the accuracy needs to be improved with more documents to train and tune up the model, addressing in particular larger documents.

Conclusions

This work used the Room Theory created by Lipizzi et al. (2021) to frame the solution to identifying relevant and important documents of varying lengths and specific parts of documents. These documents are specific to the domain of recommendations for the DAU of the DoD to implement. This approach used 3.1 GB of documents to train a vectorization model to create a vocabulary of 300-dimension word vectors. The documents were compared for similarity to the benchmark keywords and phrases on a word-to-word level. Moving average similarities were calculated to highlight the relevant/important parts of the documents for review without skimming the whole text. For evaluation, we used 10 recommendation documents and one control document, where the 10 documents scored relatively high while the control scored poorly as expected.

For future improvement, the documents could be broken up into sentences as a whole document doesn't always have the same central point. Once broken up, the sentences would be clustered together by similarity to form more cohesive content, creating logical paragraphs. These clusters/paragraphs will then be used as documents in this room theory implementation. This will help combat the issue of low-scoring similarities for lengthy documents. We already developed a prototype for this implementation.

References

D'Andrea, E., Ducange, P., Bechini, A., Renda, A., & Marcelloni, F. (2019). Monitoring the public opinion about the vaccination topic from tweets analysis. *Expert Systems with Applications, 116*, 209–226.

Gourlarte, B. F., Nassar, S. M., Fileto, R., & Saggion, H. (2019). A text summarization method based on fuzzy rules and applicable to automated assessment. *Expert Systems With Applications, 115*, 264–275.

Jimenez-Marques, J., Gonzalez-Carrasco, I., Lopez-Cuadrado, J., & Ruiz-Mezcua, B. (2019). Towards a big data framework for analyzing social media content. *International Journal of Information Management, 44*, 1–12.

Li , Y., Feng, A., Li, J., Mumick, S., Halevy, A., Li, V., & Tan, W.-C. (2019). Subjective Databases. *Proceedings of the VLDB Endowment*, *12*(11), 1330–1343.

Lin, W.-H., Wilson, T., Wiebe, J., & Hauptmann, A. (2006, June). Which side are you on? Identifying perspectives at the document and sentence levels. *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL-X)*, 109–116.

Lipizzi, C., Borrelli, D., Capela, F. de O. (2021, November 20). *A computational model implementing subjectivity with the "Room Theory". The case of detecting emotion from text* (arXiv:200506059). http://arxiv.org/abs/2005.06059

Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013, September 6). *Efficient estimation of word representations in vector space* (arXiv:13013781). http://arxiv.org/abs/1301.3781

Minsky, M. (1974). *A framework for representing knowledge*.

Pinto, J. P., & Murari, V. (2019, April). Real time sentiment analysis of political twitter data using machine learning approach. *International Research Journal of Engineering and Technology, 6*(4), 4124.

PyMuPDF. (n.d.). PyPI. https://pypi.org/project/PyMuPDF/

Wu, M., Li, Q., Wang, S., & Hou, J. (2019, August). A subjectivity-aware algorithm for label aggregation in crowdsourcing. *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, 373–378.

PANEL 21. RETHINKING EARNED VALUE MANAGEMENT IN INCREMENTAL ACQUISITIONS

Thursday, May 12, 2022	
12:30 p.m. – 1:45 p.m.	**Chair: Lieutenant General David G. Bassett, USA,** Director, Defense Contract Management Agency ***Rethinking Integrated Program Management in Incremental Acquisitions – Third-Party EVMS Compliance Assessment Study*** Gordon Kranz, Acquisition Innovation Research Center ***Improving Software Cost Estimating Techniques in Defense Programs*** First Lieutenant Nickolas A. Biancalana, Air Force Institute of Technology Jonathan D. Ritschel, Air Force Institute of Technology Lieutenant Colonel Scott T. Drylie, Air Force Institute of Technology R. David Fass, Air Force Institute of Technology Edward D. White, Air Force Institute of Technology

Lieutenant General David G. Bassett, USA—is the director of the Defense Contract Management Agency, headquartered at Fort Lee, Virginia. As the director, he leads a Department of Defense agency consisting of more than 12,000 civilians and military personnel who manage more than 300,000 contracts, performed at more than 10,000 locations worldwide, with a total value in excess of $5 trillion.

Bassett assumed leadership of DCMA on June 4, 2020. He came to the agency after serving as Program Executive Officer for Command, Control and Communications-Tactical (PEO C3T) since January 2018, where he was responsible for the development, acquisition, fielding and support of the Army's tactical network, a critical modernization priority.

Bassett was commissioned into the Signal Corps in 1988 through ROTC concurrent with a Bachelor of Science in Electrical Engineering from the University of Virginia. As a junior officer, he served in Germany in tactical positions with the 2nd Armored Cavalry Regiment and 123rd Signal Battalion, 3rd Infantry Division.

Following the Signal Officer's Advanced Course and completion of a Master of Science in Computer Science through the University of Virginia, Bassett was assigned to the U.S. European Command Staff, where he served as the Requirements Analysis and Interoperability Action Officer, J6.

He transferred to the Army Acquisition Corps in 1999 and was assigned to Fort Monmouth, New Jersey, as Operations Officer, Communications and Electronics Command Software Engineering Center. Bassett went on to manage software development efforts for the Army's Future Combat Systems program. He then served on the Joint Staff as the Ground Maneuver Analyst, Capabilities and Acquisition Division, J8.

From July 2009 to May 2012, Bassett served as the Army's Project Manager for Tactical Vehicles within the Program Executive Office for Combat Support & Combat Service Support (PEO CS&CSS). He then managed the Joint Program Office, Joint Light Tactical Vehicles (JLTV), through the Engineering and Manufacturing Development award.

In September 2013, Bassett was appointed Program Executive Officer, Ground Combat Systems, where he managed the portfolio of the Army's combat vehicle fleet including major modernization efforts to Abrams, Bradley, Stryker and self-propelled howitzer programs while also initiating the Army's Armored-Multi Purpose Vehicle program. Previous he served as Deputy Program Executive Officer for CS&CSS.

Bassett is a graduate of the Army Command and General Staff College at Fort Leavenworth, Kansas, and a distinguished graduate of the Industrial College of the Armed Forces in Washington, D.C.

Rethinking Integrated Program Management in Incremental Acquisitions – Third-Party EVMS Compliance Assessment Study

Gordan Kranz—Acquisition Innovation Research Center. [gmkranz@eipm-llc.com]

Abstract

The Department of Defense is mandating and implementing acquisition practices grounded in Agile methods to include DevSecOps. This approach allows for the incremental implementation of a system instead of fully specifying the performance. Managing complex systems requires an integrated approach that balances the technical, cost, and schedule with the end-user need. A program management team can use various tools to plan, track progress, forecast, and replan to keep the project moving forward. Earned Value Management is one of the many tools that support the program manager. It defines a disciplined set of steps for integrated planning and technical, cost, and schedule analyses. When Agile methods are implemented correctly, they meet earned value management's intent and give the program manager a continuous planning and execution process to communicate project health and status transparently in real-time. One of the aspects of earned value management is that a contractor must have an Earned Value Management System (EVMS) that meets all the requirements of EIA- STD-748D, "Earned Value Management Systems (EVMS)" (n.d.). This paper discusses an approach for an independent Third Party to assess a contractor's compliance with this standard.

What is an EVMS?

An EVMS integrates a contractor's internal business systems, processes, procedures, and tools to create a baseline plan that forms the basis for tracking progress. The business systems and processes related to EVMS consist of the Cost Accounting System (CAS), the Material Management and Accounting System (MMAS), the Cost Estimating System, the Purchasing System, the program management, and engineering cost and schedule planning processes. Figure 1 shows a simplified representation of the EVMS data integration of the planning, financial, technical, and reporting processes.

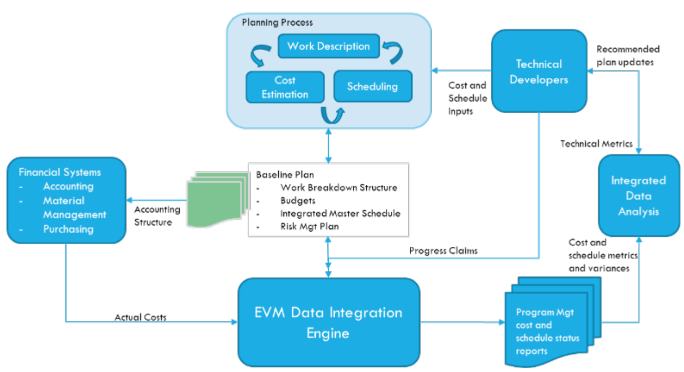


Figure 1. Simplified EVMS Representation

As one can imagine, integrating these activities into a cohesive strategy can be difficult if pre-established governance is not documented and followed. Thus, the government performs a compliance assessment to approve or disapprove a business system to ensure EVMS principles are met.

The Compliance Process

The compliance process assesses how the contractor's EVMS will integrate with various business systems and procedures to help manage the program. In addition, the compliance process will determine how the contractor will develop the technical plan, provide objective measures for monitoring progress against the plan, and report and analyze the project to support program forecasting and program management decision making.

Figure 2 provides an overview of the current DoD EVMS compliance process. The diagram shows that the EVMS requirements are manifest in many regulations, standards, and guidelines; a list of a few of those requirements follows:

1) The EIA-STD-748D Earned Value Management System Standard describes 32 guidelines that define the features of an EVMS ("Earned Value Management Systems [EIA-STD-748D]," n.d.). As with any standard, additional clarification is necessary for implementation.

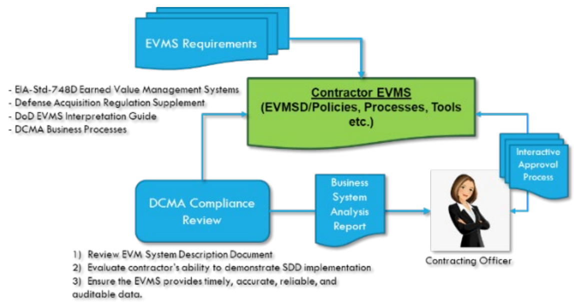


Figure 2. Overview of Current EVMS Compliance Process

2) The Department of Defense Earned Value Management System Interpretation Guide (EVMSIG; n.d.).
3) The Defense Federal Acquisition Regulation Supplement (DFARS) 252.234-7001, Notice of Earned Value Management System (n.d.); DFARS 252.234-7002, Earned Value Management System (n.d.); DFARS 252.242-7005, Contractor Business Systems (n.d.).
4) DCMA Business Processes (n.d.).

The EVMSIG is the DoD policy for EVMS and is used by the Defense Contract Management Agency's (DCMA) Earned Value Management Center to assess compliance. The compliance assessment is documented in a Business System Analysis Report (BSAS) and provided to the contracting officer, who has the authority to approve the EVMS. The initial report may note material weaknesses of the system, and if so, the contracting officer works with the contractor to resolve these deficiencies. Once the contractor addresses all the material weaknesses, the contracting officer approves the EVMS for life. However, the DCMA still conducts annual compliance surveillance to monitor the contractor's EVMS to ensure it stays compliant. For example, suppose the DCMA EVMS Center finds any material weaknesses during this yearly process. Then they can recommend to the contracting officer a review for cause or disapproval of the system. In either case, the contracting officer can require a compliance review.

The Law

On January 7, 2011, the Ike Skelton National Defense Authorization Act (NDAA) for Fiscal Year (FY) 2011 was signed into law (Pub. L. 111–383). As per the NDAA (n.d.),

"Sec. 893 Directs the Secretary to develop a program for the improvement of contractor business systems to ensure that such systems provide timely, reliable information for the management of DOD programs by the contractor and by DOD. Provides for DOD approval or disapproval of such a business system, with appropriate corrective action for disapproved systems."

There are six business systems: the Cost Accounting System (CAS), the Purchasing System, the Estimating System, the Material Management and Accounting System (MMAS), the Property Management System, and the Earned Value Management System (EVMS). To do business with the government, all contractors must have all business systems in place and approved by the government, except for EVMS. A contractor EVMS is only audited if it is performing on a development program with a cost-type contract that exceeds $100 million over the acquisition life cycle, thus significantly reducing the number of contractors who must be audited for a compliant EVMS. EVMS is also unique from the other business systems because it is primarily used to facilitate program management and less so on the auditable financial accuracy of the other methods. Keep in mind that a significant amount of taxpayers' money is being spent to execute development programs less than $100 million over the acquisition life cycle.

The FY2017 NDAA (Sec. 893) was updated and, as summarized by the NDAA (n.d.), "Requires DOD to identify and make public clear business system requirements, allow contractors to submit certifications from their third-party independent auditors (Specifically Registered Public Accounting Firms (RPAF)) that their business systems conform to DOD's business system requirements," without further review by the secretary of defense. However, the law "allows a milestone decision authority to require further auditing of business systems to manage contractual risk."

We believe the impetus of the law is primarily to have third-party auditors help the DCAA workforce get through the business system approval process. Although the DCMA is not in the same understaffed situation as the DCAA for EVMS reviews, they are not able to work with program offices to help them understand how to use the EVM data for predictive analysis. In addition, due to policy, the DCMA only reviews systems that contractors are using for development contracts over $100 million.

The Study

The (acting) deputy assistant secretary of defense for acquisition enablers and the U.S. Space Force jointly sponsored a study investigating the 2017 NDAA Section 893 concerning the earned value management business system. The investigation started in September 2021 and has a period of performance of 10 months.

The study is being done by the Systems Engineering Research Center (SERC) from Stevens University, Hoboken, NJ, as part of the Acquisition Innovation Research Center (AIRC). Key participants in the study include earned value management experts at DCMA, the Air Force, the Navy, the Space Force, the National Reconnaissance Office (NRO), NASA, and the Missile Defense Agency (MDA). Other key stakeholders helping to guide the direction of the study include OSD/ADA, OSD/DPC, the U.S. Space Force, and the Office of Management and Budget (OMB).

The technical approach of the study is to use research, interviews, and outreach for data gathering and feedback on emerging concepts.

The study's objective is to identify a model for Third-Party Assessment to increase the efficiency of EVMS approval processes and maximize the level of oversight and adherence to the specified EVMS standard.

Current Status

After agreeing on the problem description and the study objectives with the core team, we identified three types of organizations we needed to include in our conversations:

1) Consulting Firms doing EVMS compliance gap analysis
2) Defense contractors who are currently using EVMS on their programs
3) Registered Public Accounting Firms

Consulting Firms (Vendors)

We interviewed several consulting firms that defense contractors hire to conduct EVMS requirements or gap analysis as part of the contractor's preparation for a DCMA compliance review. These discussions were limited to one hour and focused on three areas: a) What EVMS requirements do you use to assess compliance against? b) What EVM(S) expertise do you use for these reviews? c) What is the scope of your reviews?

EVMS Requirements Response Summary

- All those interviewed ground their reviews with the EIA-Std-748D and the publicly available government requirements.

- Although the requirements are precise for organizations familiar with EVMS, that is not the case for contractors just getting acquainted with EVMS.

- EVMS is a unique skill set that seems to be fading across government and industry; this issue poses a challenge when performing compliance assessments.

- All vendors have standard processes and reports they provide their customers.

- During their review, a few vendors emphasized how the Integrated Master Schedule (IMS) is constructed and used the NDIA Planning & Scheduling Guide (PASEG) as a basis for their review.

Team EVM Expertise Response Summary

- None of the firms interviewed were registered public accounting firms.

- Expertise of the people doing the reviews includes those with EVMS compliance and surveillance experience, those with experience using EVM on programs, and those who have scheduling experience.

- None of the vendors required any specific EVM certification. Still, all did internal training of EVM, and most had internal testing done to assess employees' skill levels in EVM and scheduling.

Scope of the Review Response Summary

- All vendors, as a minimum, use the three basic steps of compliance: 1) Review the contractor's EVMS System Description (SDD), which should document how a company uses its internal processes to meet the EVMS requirements. 2) Assess the contractor's ability to demonstrate the use of the processes and procedures documented in the SDD, including tools. 3) Review the ability of the contractor to produce timely, reliable, and accurate data from their

EVMS, which entails reviewing a sampling of actual data from programs using EVM and generating EVM reports.

Study Observation

Companies exist that would be able to perform compliance assessments to the level of detail required for a government contracting officer to decide on the approval of a contractor's EVMS. However, further investigation is necessary to ensure the review provides the level of detail required by contracting officers.

The EVMS requirements should (could) be augmented to specify what level of detail the contracting officer requires to make a final determination.

The requirement for an RPAF to perform the compliance assessment is a problem; these companies would need to either become an RPAF or partner with one.

Large Defense Contractors

The interviews with large contractors were with companies that currently have approved systems and perform on $100 million or larger government development contracts.

The purpose of the discussions was to get their opinions on the law and whether they see the EVMS compliance process changing for them.

Contractor Interview Results Summary

1) It is not clear that having the ability to go through a third party to get a compliance assessment would benefit large companies; they all have internal organizations that do independent auditing of their business systems, including EVMS.
2) These companies expressed concern about how the interaction between the third-party assessors and DCMA might be done.
3) Large companies were also concerned about whether a third-party audit's costs would be allowable.

Study Observation

The study will need to address the concerns raised by the defense contractors and the consulting firms. For now, we offer a few considerations.

A) Consider that third-party assessment as per the law would not apply to contractors currently working with an approved EVMS.

B) The study should look for opportunities to apply the law that addresses contractors not currently covered by the $100 million threshold.

Interviews of RPAFs

The study has not held any interviews with RPAFs as of early April 2022. However, plans are to conduct those interviews by the end of April.

The Model

Based on the results of the weekly meetings, the monthly stakeholder reviews, and the interviews, the team has a Draft proposed third-party assessment model that can be used to guide the continued maturation of the concept.

Figure 3 contains a process model for addressing the update to Section 893 in the 2017 NDAA. The diagram shows a defense contractor providing certified documentation stating that it attests that the contractor's business system meets the published government requirements.

The process is identical to the current compliance process except for who does the compliance assessment.

The remainder of this study will be to refine this model supported by a detailed model description and to identify the following steps to include a possible pilot.

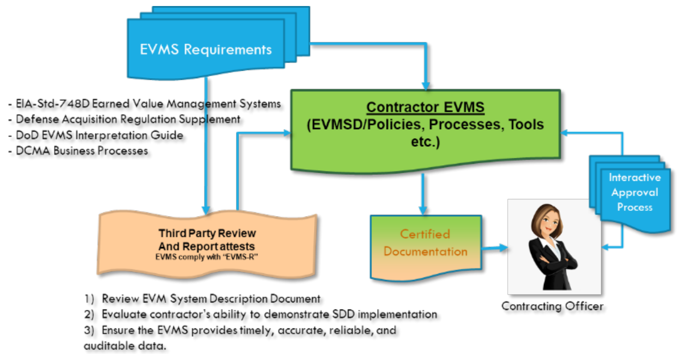


Figure 3. Proposed Third Party Assessment Model

References

DCMA business processes. (n.d.). (https://www.dcma.mil/HQ/EVMS/)

Department of Defense earned value management systems interpretation guide. (n.d.). *https://www.acq.osd.mil/asda/ae/ada/ipm/docs/DOD%20EVMIG-01-18-2019.pdf*

DFARS 252.234-7001, Notice of Earned Value Management System (n.d.).

DFARS 252.234-7002, Earned Value Management System (n.d.).

DFARS 252.242-7005, Contractor Business Systems (n.d.).

Earned value management systems (EIA-STD-748D). (n.d.). *https://www.sae.org/standards/content/eia748d/*

Ike Skelton National Defense Authorization Act for Fiscal Year 2011, H.R. 6523 (n.d.). *https://www.congress.gov/bill/111th-congress/house-bill/6523*

National Defense Authorization Act for Fiscal Year 2011, Pub. L. No 111-383 (n.d.).

National Defense Authorization Act for Fiscal Year 2017, Pub. L. No. 114-328 (n.d.).

NDIA planning & scheduling excellence guide (PASEG). (n.d.). https://www.ndia.org/-/media/sites/ndia/meetings-and-events/divisions/ipmd/links-and-reference/planning-and-scheduling-excellence-guide-paseg.ashx?la=en

Improving Software Cost Estimating Techniques in Defense Programs

First Lieutenant Nickolas A. Biancalana—is a Cost Analyst at the Air Force Cost Analysis Agency at Joint Base Andrews, MD. He received his BS in Business Management from the United States Air Force Academy and his MS in Cost Analysis from the Air Force Institute of Technology (AFIT). His research interests include strategic decision making, optimization, and data analytics. [nickolas.biancalana.1@us.af.mil]

Dr. Jonathan D. Ritschel—is an Assistant Professor of Cost Analysis in the Department of Systems Engineering and Management at AFIT. He received his BBA in Accountancy from the University of Notre Dame, his MS in Cost Analysis from AFIT, and his PhD in Economics from George Mason University. Ritschel's research interests include public choice, the effects of acquisition reforms on cost growth in DoD weapon systems, and economic institutional analysis. [jonathan.ritschel@afit.edu]

Lieutenant Colonel Scott T. Drylie, PhD—is an Assistant Professor of Cost Analysis in the Department of Systems Engineering and Management at AFIT. He received his BS in Economics from Montana State University, his MS in Cost Analysis from AFIT, and his PhD in Economics from George Mason University. His research interests include institutional economics, the economics of education, software costs, and aircraft sustainment costs. [scott.drylie@afit.edu]

Dr. R. David Fass—is an Assistant Professor of Systems Integration and Cost Analysis in the Department of Systems Engineering and Management at AFIT. He received his BA in Economics and English from the University of New Mexico, his MBA from the University of New Mexico, and his PhD in Philosophy, Business Administration & Management from New Mexico State University. Fass's research interests include cost analysis, decision analysis, and corporate and business strategies. [robert.fass@afit.edu]

Dr. Edward D. White—is a Professor of Statistics in the Department of Mathematics and Statistics at AFIT. He received his BS in Mathematics from the University of Tampa, MAS from The Ohio State University, and PhD in Statistics from Texas A&M University. His primary research interests include statistical modeling, simulation, and data analytics. [edward.white@afit.edu]

Abstract

As software becomes more ubiquitous in defense programs, there is a need to improve the accuracy and reliability of methods for estimating software size and cost. Historically, practitioners have used defined distributions in their estimating software to simulate likely outcomes. This research identifies new distributions of likely software costs and effective sizes through an analysis of Cost and Software Data Reports (CSDRs) as well as demonstrating the most appropriate distribution given certain program characteristics known at the genesis of the project. By utilizing various descriptive statistics and statistical tests, this research shows there are distributions that are more closely tailored to the actual qualities of a software program. In some instances, a broad and general distribution is sufficient; however, there are specific commodities, contractors, and system types that are distinctly different and require additional analysis. Overall, this research intends to equip practitioners with an arsenal of distributions and statistical information that will lead them to apply the best model to predict software size and cost, all with the goal of improving overall accuracy.

Introduction

Software has become a core functional element in many defense projects and therefore plays a vital role in the definition of mission critical capabilities (McQuade et al., 2019). For that reason, it is prudent to utilize a consistent and accurate method to properly quantify the expected costs associated with incorporating software into defense projects. Given the implications of improperly estimating software systems and, in turn, the entire project, (e.g., cost overruns, inadequate funding, etc.) accurate estimates are paramount. The increasingly robust centralized cost databases (e.g., Cost Assessment and Data Enterprise [CADE]) provide the opportunity for cost analysts to access a wide array of historical software data points. This data can be used to create new distributions that form accurate regions of reasonable estimates, ultimately helping the cost analyst perform a more precise estimate.

By analyzing historical defense projects from various branches in the military, this research seeks to identify patterns between different characteristics of projects and how they relate to the final cost of software packages. Once these relationships are uncovered, that information can shed light on how to properly size software to new projects. The software data from these historical projects holds the key to refining the estimating process. Additionally, it will provide practitioners with an arsenal of distributions that can be used as inputs into simulation software to create likely outcomes for the cost and size of software programs. This will increase confidence in estimates as well as provide clarity as to how software costs relate to the mission in which they are designed to serve. The implications include not only formulating more accurate estimates but also knowing what a realistic cost would be prior to accepting contractor proposals. This puts the DoD in an advantageous and leveraged position during negotiations while also mitigating potential risk of cost overruns.

Background

Considering the array of possible applications of software and the platforms in which they serve, it would be unreasonable to consider the software programs and their costs homogenous. The ability to obtain a more specific distribution of the likely costs associated with software in a project given various characteristics of the project itself is vital. The efforts of Sheppard and Schofield (1997) show that analogy methods predict software exceptionally well compared to regression-based analysis. By creating more comprehensive analogies and specific distributions, estimators will have a more refined tool to formulate accurate and precise estimates. Additionally, it will educate decision makers as to what is and is not a reasonable contractor proposal. Previous research regarding software systems in DoD programs has been conducted; however, it involved software size as it related to effort levels rather than the cost itself (Madachy et al., 2011; Sheppard & Schofield, 1997).

This research aims to explore a new and unique angle to software costs in defense projects. This research is unique in that it is looking at the individual costs of the software itself within the defense programs and using those to create comprehensive analogies to aid in future cost estimates. Up to this point, this approach has not been pursued and data had not been readily available. The Air Force Life Cycle Management Center (AFLCMC) collected and provided consolidated datasets containing not only the software characteristics of dozens of different projects, but also each project's respective cost information. This was accomplished by taking the Software Resource Data Reports (SRDR) for each project from CADE and matching those software characteristics with the cost information found on the project's Cost Data Summary Report (i.e., Form 1921). This data included information regarding the commodity, branch of service, nonrecurring costs, total lines of code, Effective Source Lines of Code (ESLOC), team structure with regards to experience level, number of hours in each phase of development and a multitude of other measurements. The dataset is among the first of its kind in that it combines a project's software data (lines of code, primary language, etc.) with its cost data, allowing for a comprehensive analysis of the relationships between cost and software as it relates to difference project characteristics.

Problem Statement/Research Questions

One of the problems this research addresses is accounting for diversity in software. There is a wide variety of defense projects, spanning an enormous range of software specifications, software requirements, and ultimately software costs. From a cost analyst's point of view, this makes estimating a software system's cost a particularly perilous task. If there were a way to narrow the range of possible values given certain characteristics of a project, the analyst would be able to provide a more accurate and confident estimate of a project's software costs. To address this problem, this paper examined the question, "How do the size and cost of software packages relate to the project in which they operate, and how do they change as the characteristics of the project are changed?

Literature Review

The motivation behind this literature was to validate or contradict the selected elements of this research. At any level of this literature review, if the elements were invalidated or found to be of little use, the purpose of the research would be of little use. Starting at the top level, software in and of itself is being increasingly relied upon in the DoD (GAO, 2021). Considering software is now at the forefront of DoD acquisition programs either in a direct or supporting role, the methods and techniques used to estimate the costs must be fortified.

Next, utilizing the results of this research implies the use of analogy and parametric estimating methods. The analogy method entails finding an analogous program and scaling its parameters to model the new program based on its known characteristics. This method has several advantages and disadvantages (Garrett, 2008; Kueng, 2008). The main disadvantages stem from the analogy itself and its appropriateness. If the analogy cannot be defended and should not be used, the estimate created has lost its value. Despite the disadvantages, this method has shown to be superior to regression-based estimating within the realm of software (Sheppard & Schofield, 1997). The parametric method involves using parametric models that have been derived from cost driving factors that are found by developing statistical relationships between historical costs and program, physical, and performance characteristics (Garrett, 2008). This method also has its advantages and disadvantages (Pfleeger et al., 2005) and has been refined through decades of research. AFLCMC uses a form of parametric modelling in their software cost estimations as well. They input known distributions for various project parameters into their estimating software and perform simulations. These simulations result in distributions for overall costs that are used in the decision-making process.

Next, regarding the independent variables of this research, previous works have segmented datasets into groups that resemble the groups used in this research. Jones et al. (2014), although investigating a common rule of thumb in O&S cost estimating, segments their dataset into groups labeled Space, Fixed-Wing Aircraft, Rotary-Wing Aircraft, Missiles, Electronics, Ships, Surface Vehicles, and Automated Information Systems (AIS). They further segment Fixed-Wing Aircraft into Fighter, Cargo/Tanker, and Unmanned Aerial Vehicles (UAV). These groups very closely resemble the commodity and system type groups used in this research. Their results showed the need to segment projects by these categories and found differences between them (Jones et al., 2014). Additionally, as part of their statistical analysis, Madachy and Clark (2015) segmented their data by "operating environment." Members of this group included Aerial Vehicle (including fixed-wing, rotary-wing, and unmanned aircraft), Space Vehicle, and Ordnance Vehicle (including missiles). Simultaneously, given the missions of each project, each article is also segmenting their datasets by Service, although not explicitly. These sources show an intuition to separate projects and create homogenous groups such as commodity and system type and explain it is unwise and imprudent to treat all projects the same.

Lastly, the use of these specific dependent variables must be validated. This research utilized Effective Source Lines of Code (ESLOC) and nonrecurring costs as a rate of ESLOC. The Air Force Cost Analysis Agency (AFCAA) and Naval Center for Cost Analysis (NCCA) describe effective size as a major factor of software cost and schedule estimating (AFCAA, 2008). AFCAA goes on to explain the role of Effective Source Lines of Code (ESLOC) and how it relates size to work. They explain,

> Resource estimates based on physical source lines of code for modified software and systems containing reusable components, cannot account for the additional resource demands attributable to reverse-engineering, test, and software integration. The common method used to account for the added resource demands is to use the *effective* software size. (AFCAA, 2008)

Additionally, Clark and Madachy further this statement in the Software Cost Estimation Metrics Manual for Defense Systems (2015) and state equivalent size is "a key element in using software size for effort estimation" (Clark & Madachy, 2015). They go on to assert that equivalent size quantifies how much effort is required to reuse old code alongside new code. ESLOC is a pivotal measurement that encapsulates both size and complexity.

With the introduction of cost, it is important to distinguish between recurring and nonrecurring costs. The Defense Acquisition University defines nonrecurring costs as "costs that will occur once or occasionally for a particular cost objective, NRCs include preliminary design effort, design engineering, and all partially completed reporting elements manufactured for tests" (DAU Glossary, n.d.). Additionally, they describe a recurring cost as "costs for items and services that reoccur, especially at regular intervals. Recurring costs are incurred each time a unit equipment is produced, such as direct labor and direct materials" (DAU Glossary, n.d.). Since the costs that this research is focused on is the preliminary design and engineering of software packages, nonrecurring costs will be assessed in the form of the rate nonrecurring cost per ESLOC.

Research Gap

Previous research has looked at past relationships between software size, effort, productivity, and complexity, but normalized historical costs have not been included in the analysis. This current research is not only aimed at utilizing previous costs to establish relationships and distributions to predict future costs but also investigating program characteristics and how they influence key cost drivers such as ESLOC.

The data from the CADE database directly links the software characteristics from a program's Software Resources Data Report to its cost data from its Form 1921. Previous research has investigated software through various lenses; however, this dataset finally allows for the direct analysis of past costs and not relying on some form of a proxy to estimate costs. This offers the opportunity to poignantly investigate distributions regarding both the cost per ESLOC parameter and the ESLOC parameter itself. As stated earlier, ESLOC is a key metric in software models that encapsulates the size and effort of a project. Additionally, creating a rate of cost per ESLOC standardizes each project in the dataset to avoid distortions from exceptionally large and/or expensive projects.

Data

The data used in this research is a combination of datasets from AFLCMC and CADE. AFLCMC provided the consolidated data containing program characteristics, software components and capabilities, nonrecurring costs, and many other quantities and dates pertaining to the development and purchase of the software packages. This dataset contained 44 different programs across the DoD with detailed information down to the WBS element. This dataset is a consolidation of the WBS element's software characteristics and properties found on the project's Software Resource Data Reports (SRDR) and the element's cost information found on the project's DD Form 1921. This data was collected by AFLCMC from CADE and consolidated for this research. The AFLCMC data was verified by taking a 10% (50 WBS elements) sample and comparing the information to the sourced data from CADE. Once the sample was taken, each WBS element from the AFLCMC data was found in the CADE dataset and compared for accuracy. Of the 50 WBS elements used, all matched the CADE dataset giving confidence there are few or no mismatches in the AFLCMC data.

Inclusion Criteria

Now that there is a consolidated and verified dataset, lines containing outliers or missing data must be excluded from the analysis. Due to the highly skewed nature of this dataset, a more traditional outlier test such as three times the Interquartile Range (IQR) beyond the 25th and 75th percentiles was not practical. The skew present in the data caused the IQR to be very small which would place the outlier bounds closer to the median. If this approach were taken, nearly 10% of this dataset would be excluded. For this research, initial outliers were identified using a quantile range exclusion method (Klimberg & McCollugh, 2016). This method calculated the range from the fifth to 95th percentiles, multiplied this range by three and excluded any data points beyond that distance from the fifth and 95th percentiles. For example, with regards to Nonrecurring Cost/ESLOC, the calculation is as follows:

$$Range = 95th\ Percentile - 5th\ Percentile \quad (1.1)$$

$$Range = \$3.767K - \$0.009K = \$3.758K$$

$$Outlier\ Lower\ Bound = 5th\ Percentile - (Range * 3) \quad (1.2)$$

$$Outlier\ Lower\ Bound = \$0.009K - \$11.274K = -\$11.265$$

$$Outlier\ Upper\ Bound = 95th\ Percentile + (Range * 3) \quad (1.3)$$

$$Outlier\ Upper\ Bound = \$3.767K + \$11.274K = \$15.041K$$

With these bounds now established, any observations beyond them were excluded from the dataset. Since the lower bound was negative and neither Cost per ESLOC nor ESLOC can be negative, the values were truncated at zero. This technique was performed for both the Cost/ESLOC analysis as well as the ESLOC analysis. For the rate analysis, three observations were removed and for the ESLOC analysis, two were removed.

Table 1. Distribution Analysis Sample Sizes with Exclusions

	Cost/ESLOC Analysis	**ESLOC Analysis**
Total Initial Data Points	460	460
Missing Values	106	66
Outliers	3	2
Data Points Remaining (all analyses except contract type)	**351**	**392**
No viable contract information	37	38
Data Points Remaining (contract type analysis only)	**314**	**354**

Methodology

The overall approach for this research was a process dubbed "incremental analysis." The purpose of incremental analysis is to observe how the dependent variables (Cost/ESLOC and ESLOC) change as other variables are changed. Traditionally, a regression model would show the individual effects of each independent variable on the dependent variable. However, there were many interactions between independent variables within this dataset that would decrease the overall utility of the model. If a regression model were pursued, the outcome would contain many interactions variables pertaining to specific combinations of contractor and commodity, commodity and service, contract type and commodity, and so on. The resulting regression model would indicate effects on the dependent variables; however, they would only apply to those specific combinations and would lack utility.

The alternative is to do a series of bivariate analyses with various combinations of independent variables to observe how the dependent variable changes. Additionally, these analyses would show which individual combinations are different from one another thus identifying variables that have more impact on the dependent variables than others. These unique differences also illuminate which combinations of independent variables require a distribution of their own outside of the univariate distributions found for each individual variable.

The incremental analysis was performed twice for each pair of independent variables. One analysis for a given pair of characteristics holds one independent variable constant while varying the other and the second analysis switches the variables. Within each analysis, the median cost ($K/ESLOC) and effective size (KESLOC) is reported. This is due to the skewed nature of the data and as a result, a mean would not be a good representation for the data. Given there are five program characteristics in this research, there are 25 total combinations. This method is repeated for each combination of independent variables except for the combination of commodity and system type since system type is a subgroup of commodity. Additionally, a characteristic will not be compared against itself. After these removals, there were 18 total combinations explored in this research. All 18 combinations will be explored for both dependent variables, Cost per ESLOC and ESLOC. For the purposes of this paper, only those pertaining to a project's commodity are discussed.

Each analysis contains a Kruskal-Wallis p-value which compares the values within the constant variable as it's changed by the other variable. This p-value indicates whether differences are detected between the values and the subsequent Steel-Dwass test highlights which specific pairs of values are different from one another. An alpha level of 0.05 was used for both tests. The Steel-Dwass outcomes for each pair of analyses are then compared and any overlaps in results indicate a specific combination of variables that warrants its own distribution. This is because two Steel-Dwass tests have shown that each variable that is part of the specific combination was different than at least one of the other categories within its subset.

These specific combinations were then fit with multiple probability density functions (PDFs), and each was evaluated on how well it fit the distribution. Due to the practitioner's familiarity, lognormal distributions were always provided, regardless of whether it was the best fit or not. The Anderson-Darling test result is provided so the practitioner is aware if a lognormal distribution is not an appropriate method to model this data and should use the better fitting distribution.

Analysis and Results

This section contains incremental analyses showing how Cost/ESLOC and ESLOC change when one variable varies and another is held constant, all in search of more specific combinations of variables that warrant a separate distribution.

Cost Per ESLOC Analyses

The following analysis identifies how Cost/ESLOC changes as various independent variables are changed. Each iteration of this analysis will take two independent variables, hold one constant, and assess how the median values of Cost/ESLOC change as the Other independent variable is changed. The variables are then switched regarding which is held constant to identify any unique pairs of variables that warrant a deeper analysis.

Cost/ESLOC—Commodity and Contractor

Table 2 illustrates a two-way dissection of the Cost/ESLOC rate. It segments the data first by contractor, then by commodity. It also shows the differences between commodities within the same contractor. The numbers within the table represent the median Cost/ESLOC in thousands for each commodity within each contractor. The bottom three rows of the table show the total number of observations, median value, Kruskal-Wallis p-value for the test performed on the commodities within a certain contractor. Steel-Dwass pairs are annotated by shared letters in the cells.

Table 1. Contractor by Commodity Analysis—Cost ($K)/ESLOC

Commodity/Contractor	Contractor 1	Contractor 2	Contractor 3	Contractor 4	Contractor 5
Aircraft	0.183	0.185	0.174	0.088[b]	0.043[b]
EAS	0.278	0.177[a]	0.056	0.134[a]	0.071[a]
Missile		0.708			0.511[a,b,c]
Rotary Wing	0.202	0.467	0.218	0.327	
Space	0.798	1.331[a]	0.131	1.051[a,b]	0.078
UAV			0.173		0.141[c]
N	96	46	89	52	68
Median	0.205	0.377	0.119	0.28	0.122
KW	0.303	0.004	0.05	0.001	0.008

Note: Commodities that share a letter within the same contractor are members of a Steel-Dwass Pair

The correct way to interpret Table 2 is as follows. The Kruskal-Wallis p-value (0.303) is not smaller than 0.05, meaning there is not sufficient evidence to say any of the commodities within Contractor 1 are different from one another.

The same results are found when looking at Contractor 3 in that none of the commodities are distinctly different. However, Contractors 2, 4, and 5 all have significantly low Kruskal-Wallis p-values, indicating that there are differences between commodities within the single contractors. Additionally, one can determine which contractor produces more expensive commodities by comparing the median values of each member of a Steel-Dwass pair.

The results of this analysis show that even within a singular contractor, differences can be found between commodities. Additionally, these results show that while holding contractor constant, the Cost/ESLOC changes as commodity changes. Prior to this analysis, the contractor or commodity was analyzed at large but this shows that even within a particular contractor, further analysis may still be required to find the most appropriate distribution.

This analysis only represents one side of this investigation. Although this analysis showed differences within a specific contractor, if a difference cannot be found between contractors within the same commodity, then one would be better off to use the overall commodity distribution. However, if a difference is found between commodities within the same contractor and that same difference is found between contractors within the same commodity, an even more specific distribution would be required.

Table 3. Commodity by Contractor Analysis—Cost ($K)/ESLOC

Contractor/Commodity	Aircraft	EAS	Missile	RW	Space	UAV
Contractor 1	0.183	0.278[a,b]		0.202	0.798	
Contractor 2	0.185	0.177	0.708	0.467	1.331[b]	
Contractor 3	0.174	0.056[a]		0.218	0.131[a,b]	0.173
Contractor 4	0.088	0.134		0.327	1.051[a]	
Contractor 5	0.043	0.071[b]	0.511		0.078	0.141
N	75	105	21	61	38	51
Median	0.149	0.127	0.577	0.207	0.57	0.143
KW	0.308	0.001	0.622	0.334	0.005	0.434

Note: Contractors that share a letter within the same commodity are members of a Steel-Dwass Pa

Table 3 shows the same median values, but this analysis switches the rows and columns compared to the previous analysis. This analysis now shows the differences between contractors within the same commodity. Again, the Kruskal-Wallis p-values show whether significant differences were detected within a given commodity. This test found significance in the EAS and Space commodities. Contractors that share a letter within the same commodity are members of a Steel-Dwass pair and their relationship to one another can be found my comparing median values.

Like the previous one, this analysis identifies differences between contractors within the same commodity as well as which produces a more or less expensive software component than another. This shows that segmenting the data by only commodity may still not be sufficient given the differences that were found.

If there are overlapping differences in the preceding analyses, additional distributions are required. For example, looking at the Steel-Dwass pairs from the Contractor by Commodity analysis, Contractor 2 has a Steel-Dwass pair of Space and EAS. This means that within Contractor 2, Space and EAS are distinctly different from one another. Knowing this information and looking at the Commodity by Contractor analysis, if Contractor 2 is a member of a Steel-Dwass pair for either Space or EAS, that would require a new distribution since both components have been shown to be distinctly different. This instance occurs with Contractor 2 Space programs. Simply put, Contractor 2 Space programs have shown to be different than other commodities that Contractor 2 works on and different than space programs that other contractors do. This warrants an additional distribution due to the dual differences found. This distribution is shown below.

Table 4. Contractor 2/Space Distributions—Cost ($K)/ESLOC

	Median	Distribution	AD
Contractor 2 - Space	1.331	Exponential	0.968
		Lognormal	0.879

Table 4 shows the distribution information for all data points representing Space programs accomplished by Contractor 2. An Exponential distribution is the best fit based on AIC and the p-value indicates that it is appropriate to use an exponential distribution to model this information. Additionally, a lognormal distribution would also be appropriate given the parameters above. Since this dataset is small it is difficult to draw firm conclusions regarding the overall shape of these data points but given the data at hand, these distributions would be appropriate. This phenomenon occurs twice more, the first being Contractor 4 (mostly subcontractors or contractors with few data points) and Space.

Table 5. Contractor 4/Space Distributions—Cost ($K)/ESLOC

	Median	Distribution	AD
Contractor 4 - Space	1.051	Exponential	0.758
		Lognormal	0.987

Like the previous distribution, both an exponential and lognormal distribution would be an appropriate fit. Again, due to a small sample size, it is difficult to draw firm conclusions; however, based upon the data at hand, these distributions would be appropriate. Lastly, this occurs again with Contractor 5 and EAS.

Table 6. Contractor 5/EAS Distributions—Cost ($K)/ESLOC

	Median	Distribution	AD
Contractor 5 - EAS	0.071	Exponential	0.179
		Lognormal	0.293

These three distributions represent the most specific and detailed level that retains relevancy. Other than these, the lowest level required would be either the contractor level or commodity level. However, since the components of these three distributions have shown to be different than their counterparts', more detailed distributions are required. The remaining analyses were performed in the same manner as outlined above but their results are presented in an abridged format.

Cost/ESLOC—Commodity and Service

The next analysis held commodity constant while varying service. After performing the statistical tests, it was found that only the EAS commodity had significant differences between services. The Army was found to be significantly cheaper than both the Air Force and Navy.

Next, commodity was varied within each service. Both the Air Force and Army had significant differences detected. Within the Air Force, Space was found to be more expensive than both Aircraft and EAS. Within the Army, EAS was found to be cheaper than Missile and Rotary Wing. These results coupled with the results of the previous analysis show there are two instances where a more specific distribution is required.

Table 7. Air Force/EAS Distribution—Cost ($K)/ESLOC

	Median	Distribution	AD
Air Force - EAS	0.192	Lognormal	0.764

The table above shows the best distribution for the specific combination of Air Force and EAS. Based on AIC, a lognormal PDF was the best fit for this distribution and based on its Anderson-Darling p-value, it is also an appropriate method to model data with these characteristics.

Table 8. Army/EAS Distribution—Cost ($K)/ESLOC

	Median	Distribution	AD
Army - EAS	0.057	Lognormal	0.523

The other overlap occurs again with EAS but this time with the Army. Again, lognormal was the best fit based on AIC and is appropriate based on Anderson-Darling p-value.

Cost/ESLOC—Commodity and Contract Type

The next analysis investigates how Cost/ESLOC changes when contract type is held constant, and commodity is changed. None of the specific contract types contained significant differences between commodities except for Mixed Contracts (MC). Space was found to be more expensive than Rotary Wing, EAS, Aircraft, and UAV. Given there is no clean definition as to what exactly comprises a Mixed Contract, it is difficult to draw any firm conclusions from these results.

Next, commodity and contract type were switched to investigate how Cost/ESLOC changed when commodity is held constant and contract type is varied. Only Aircraft and EAS had detectable differences. Within Aircraft, CPFF contracts were cheaper than Mixed Contracts. Within EAS, CPFF was found to be cheaper than CPAF, CPIF, CW, and MC.

There are two specific combinations of commodity and contract type that appears in both sets of Steel-Dwass pairs and warrants a more specific distribution. The first distribution is shown below.

Table 9. MC/EAS Distribution—Cost ($K)/ESLOC

	Median	Distribution	AD
MC - EAS	0.079	Lognormal	0.152

Table 9 shows the best fitting PDF for EAS commodity WBS elements performed on a Mixed Contract. Based on AIC, lognormal was the best fit and based on the Anderson-Darling p-value, it is also an appropriate fit due to the value being larger than the alpha level of 0.05. It is difficult to put the utility of this distribution in perspective since there is no clear definition of a Mixed Contract in terms of composition of fixed versus cost-plus elements.

Table 10. MC/Aircraft Distributions—Cost ($K)/ESLOC

	Median	Distribution	AD
MC - Aircraft	0.259	Exponential	0.372
		Lognormal	0.026

Table 10 shows the best fitting distribution for Aircraft commodities using a mixed contract type. Exponential was the best fit based on AIC and the Anderson-Darling p-value shows it is appropriate to use. Lognormal is provided but the p-value shows it is not an appropriate PDF to use to model this data.

ESLOC Analyses

The following analyses are identical in nature to the Cost/ESLOC analyses shown previously except now the dependent variable is ESLOC in thousands. These results will show how ESLOC changes when one independent variable is held constant and the Other is changing, illuminating the impacts of these independent variables.

ESLOC—Commodity and Contractor

Like the previous contractor by commodity analysis performed on Cost/ESLOC, this analysis showed how ESLOC changes when the commodity is changed, all while contractor is held constant. Only Contractors 1 and 2 had detectable differences between commodities. Within Contractor 1, Aircraft projects showed to lead to a significantly larger effective size than Rotary Wing. Within Contractor 2, Missile was found to be smaller than EAS and Space.

Next, commodity was held constant while contractor was varied. Only within the Missile commodity were differences found. Contractor 2 was found to create much smaller software programs than Contractor 5.

One overlap occurred with Contractor 2's Missile projects. The combination of Contractor 2's Missile projects being different than other commodities in which they have performed work and different than other contractors' Missile projects warrant a separate distribution to model this specific relationship.

Table 11. Contractor 2/Missile Distributions—ESLOC (K)

	Median	Distribution	AD
Contractor 2 - Missile	4.948	Exponential	0.694
		Lognormal	0.004

The exponential distribution was the best fit for this data based on AIC and is appropriate due to the Anderson-Darling p-value. The lognormal distribution on the other hand is not an appropriate tool to model this data since its p-value is below the alpha level of 0.05.

ESLOC—Commodity and Service

The next two analyses investigate the impact on ESLOC (K) when commodity and service are changed. The first changes service while holding commodity constant. Only within the EAS, Missile, and UAV commodities were differences detected. Within the EAS commodity, yielded significantly larger effective sizes than both the Air Force and Navy. Within the Missile commodity, The Army was significantly smaller than the Air Force. Lastly, within the UAV commodity, the Navy was significantly larger than the Air Force.

Next, the previous independent variables are switched, and service is held constant while commodity is varied. Within the Army, Rotary Wing was found to be larger than Missile, and EAS was found to be larger than both Missile and Rotary Wing. Within the Navy, EAS was found to be smaller than Aircraft.

There are three instances where the Steel-Dwass pairs overlap and require a more specific distribution.

Table 12. Army/Missile Distributions—ESLOC (K)

	Median	Distribution	AD
Army - Missile	4.948	Exponential	0.649
		Lognormal	0.006

The first overlap occurs with Missile WBS elements performed by the Army. The table above shows exponential as the best fitting PDF based on AIC. Due to the practitioner's familiarity with lognormal distributions, it is also provided. Based on Anderson-Darling p-values, the exponential distribution is an appropriate method to model this data since the p-value is larger than the alpha level of 0.05. However, a lognormal distribution is not an appropriate method because the p-value is less than the alpha level.

Table 13. Army/EAS Distribution—ESLOC (K)

	Median	Distribution	AD
Army - EAS	214.071	Lognormal	0.195

The next overlap occurs with the Army and the EAS commodity. Lognormal was the best fitting PDF based on AIC and an appropriate model based on the Anderson-Darling p-value.

Table 14. Navy/EAS Distribution—ESLOC (K)

	Median	Distribution	AD
Navy - EAS	53.56	Lognormal	0.291

The final overlap occurs with Navy/EAS projects. For these projects, a lognormal distribution is the best PDF to use and is also appropriate based on the Anderson-Darling p-value.

ESLOC—Commodity and Contract Type

This analysis investigates the impact on ESLOC as contract type and commodity are changed. Like the Cost/ESLOC sections regarding contract type, since not all data points had viable contract information, some needed to be scrubbed from the dataset. The first analysis holds contract type constant while varying commodity. Three different contract types had detected differences between commodities. Within CPAF, Rotary Wing was found to be smaller than Aircraft. Within CPIF, EAS was found to be larger than Aircraft. Lastly, within MC, UAV was found to be smaller than Space.

Next, commodity was held constant while varying contract type to see the impacts on ESLOC. Within the EAS commodity, CPFF was found to be larger than CPAF, CPIF, and MC. Within the UAV commodity, CPIF was found to be larger than MC.

Overlaps in the two iterations of Steel-Dwass test highlight which specific combinations of commodity and contract type warrant a more specific distribution. Two such overlaps occur, and each distribution is shown in the following tables.

Table 15. EAS/CPFF Distributions—ESLOC (K)

	Median	Distribution	AD
EAS - CPFF	475.791	Exponential	0.052
		Lognormal	0.057

Table 15 reflects the distribution parameters that form a PDF modeling the data for EAS commodities performed with a Cost-Plus Fixed Fee contract. Based on AIC, exponential is the best fitting distribution but lognormal is also provided due to the practitioner’s familiarity. Both distributions are appropriate methods to model this data given that both Anderson-Darling p-values are larger than the alpha level of 0.05.

Table 16. UAV/MC Distributions—ESLOC (K)

	Median	Distribution	AD
UAV - MC	25	Gamma	0.8
		Lognormal	0.645

The last overlap occurs for UAV elements performed under a Mixed Contract. Based on AIC, a Gamma distribution was the best fit and both it and a lognormal distribution would be appropriate means to model this specific data since both Anderson-Darling p-values are larger than the alpha level of 0.05. As in previous distributions pertaining to Mixed Contracts, the utility is difficult to define since Mixed Contracts can vary drastically. There is no clear definition of a Mixed Contract Other than possessing fixed and cost-plus elements. The proportions, however, are not defined.

Results, Limitations, and Future Research

This research was oriented toward identifying the various distributions that can be used to model the values of Cost per ESLOC and ESLOC within software programs. Segmenting the dataset by different program characteristics (e.g., service, commodity, contractor, and contract type) highlighted elements of a project that can influence the size and cost of software in defense programs. Additionally, by incrementally changing various characteristics, one can see the marginal changes in each dependent variable as a certain project element is varied.

Results

The findings from this research emphasize the heterogeneity found in Cost per ESLOC and ESLOC values. Although overall distributions can be used to model these values, the results shown earlier indicate that certain characteristics of a project can change the region of plausible values and can aid in creating more specific distributions. The results show that some contractors, commodities, services, and contract types tend to result in bigger or more expensive program elements. Knowing this, it may not always be advisable to use a general distribution when a more specific one is available.

The incremental analyses showed how Cost per ESLOC and ESLOC changed within certain program characteristics. The incremental analyses served the same purpose as a linear regression in that it analyzed how a dependent variable (Cost/ESLOC, ESLOC) varied when another is held constant. Put another way, it showed the marginal changes in the dependent variable because of a change in an independent variable. Each pair of analyses (those with the same independent variables but the one held constant and the one varied were switched) were compared and when overlaps in Steel-Dwass pairs were present, this highlighted the need for a more specific distribution, one tailored to a particular pair of characteristics.

The following flowcharts (see Figure 1 and Figure 2) provide a roadmap for the practitioner to arrive at a recommended distribution to use in their software cost model. The practitioner starts at the left side with the commodity of interest. Moving to the right, one enters another characteristic of the program, in this case it is service. If any of the conditions are met within service, the distribution identifier is provided, and the practitioner stops. (Note: Readers can contact the authors for specific distributions.) If no conditions are met in the service section, the user moves to the next section. Once the user has moved through the entire flowchart, if no intermediate conditions have been met, the identifier for the overall distribution of that commodity should be used. If multiple distributions apply to a given project, any of them can be used to model outcomes; however, it will be at the practitioner's discretion to determine which to use.

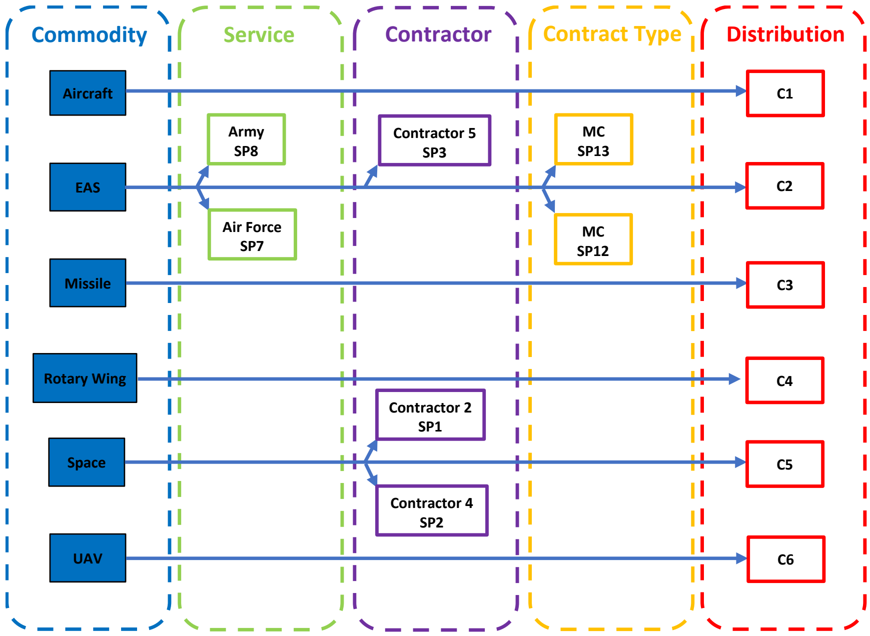


Figure 1. Cost/ESLOC Distribution Flowchart

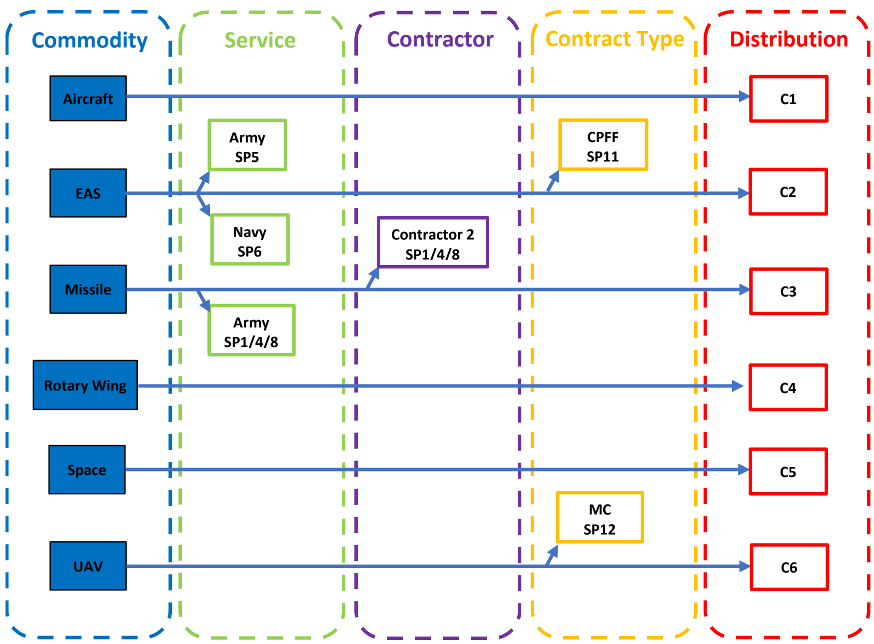


Figure 2. ESLOC Distribution Flowchart

The flowchart could have begun with any of the characteristics; however, commodity was chosen for the following reasons. If this were to be performed very early in the decision-making process, contractor and contract type may not be known. System Type was not used because it is a subset of commodity, subjective in nature, and in many instances, it simply mirrors the results of commodity. Lastly, service was not used since this research was primarily intended for use at AFLCMC, an Air Force entity. For their purposes, they are interested in Air Force programs and if service was the origin of the flowchart, some specific distributions would be left out, perhaps distributions that would better fit the program of interest.

Limitations

Limitations to this research are mostly related to the dataset. Regarding the various program characteristics, some do not have a well-defined definition and therefore introduce subjectivity. "Mixed contract" does not have a clear definition outlining the proportion of fixed and cost-plus elements. This means that two contracts with wildly different proportions could both be considered a mixed contract and would therefore utilize the same distribution. There were differences found between certain contract types so it would be beneficial to know proportions of fixed and cost-plus elements.

There were also limitations regarding the process used to obtain results. As mentioned before, a traditional regression analysis could not be performed due to overlaps and interaction found in the dataset. For this reason, the incremental approach was taken and although it is a rather laborious substitute, the rationale is largely the same. Since it was not a regression analysis, coefficients were not calculated and thus, firm conclusions regarding a particular characteristic's impact could not be illustrated, only direction.

Lastly, some projects had more lines in the dataset than others meaning it was more represented in each characteristic. As a result, some commodities, contractors, etc. had more data points not because there were more projects but because there were more WBS elements.

Future Research

If more data can be collected and utilized for these purposes, other methods could be employed in future research. A conventional regression analysis could be performed, and the coefficients would indicate the true impact on Cost per ESLOC and ESLOC. In theory, a regression equation could be formulated to predict the size and cost of a program given only characteristics, a tool that could prove to be invaluable to cost estimators and high-level decision makers alike.

One area that is ripe for future research involves team productivity. One could utilize the cost data in this research and analyze it as a rate of dollars spent per manhour on the project. Also, they could move toward efficiency and investigate hours/ESLOC. Both rates could be analyzed through each characteristic to expose differences and highlight in which situations software development teams tend to be more productive or efficient.

This research illuminates the patterns which costs, and effective sizes follow with regards to various elements of a software program. With these software cost and size distributions, a practitioner can pick the distribution that applies the project they are estimating and know it was created for that exact situation. This research serves as a first step in identifying distributions between software program elements and the costs that are incurred as a result, all with the intent to increase the overall accuracy and effectiveness of cost estimation.

References

Air Force Cost Analysis Agency & Naval Center for Cost Analysis. (2008). *Software development cost estimating handbook.* https://www.dau.edu/cop/ce/dau%20sponsored%20documents/sw%20cost%20est%20manual%20vol%20i%20rev%2010.pdf

Clark, B., & Madachy, R. (2015). Software cost estimation metrics manual for defense systems. *Software Metrics Inc.*

Garrett, G.A. (2008). Cost estimating methods and Sources of risk. *Contract Management, 48*(10), 14–24.

Jones, G., White, E., Ryan, E. T., & Ritschel, J. D. (2014). *Investigation into the ratio of operating and support costs to life-cycle costs for DoD weapon systems*. Air Force Institute of Technology Wright-Patterson AFB OH.

Keung, J. (2009). Software development cost estimation using analogy: A review. *2009 Australian Software Engineering Conference.* https://doi.org/10.1109/aswec.2009.32

Klimberg, R., & McCullough, B. D. (2016). *Fundamentals of predictive analysis with Jmp* (2nd ed.). SAS Institute Inc.

McQuade, J. M., Murray, R. M., Louie, G., Medin, M., Pahlka, J., & Stephens, T. (2019, March 12). *Periodic table of Devops tools V3 PDF*. Retrieved July 28, 2021, from https://media.defense.gov/2019/Mar/14/2002101480/-1/-1/0/DIB-SWAP_STUDY_REPORT[DRAFT],_LAST%20MODIFIED_13MAR2019.PDF.

Pfleeger, S. L., Wu, F., & Lewis, R. (2005). *Software cost estimation and sizing methods: Issues, and guidelines.* Rand Corp.

Shepperd, M., & Schofield, C. (1997). Estimating software project effort using analogies. *IEEE Transactions on Software Engineering, 23*(11), 736–743. https://doi.org/10.1109/32.637387

PANEL 22. ENSURING CYBERSECURITY ACROSS THE ACQUISITION ECOSYSTEM

Thursday, May 12, 2022	
12:30 p.m. – 1:45 p.m.	**Chair: Rear Admiral Kurt Rothenhaus, USN,** Program Executive Officer, Command, Control, Communications, Computers and Intelligence ***Defensive Industrial Policy: Cybersecurity Interventions to Reduce Intellectual Property Theft*** Chad Dacus, Air Force Cyber College Carl (Cj) Horn, Air Force Cyber College ***Digital Engineering Effectiveness*** Alfred Schenker, Carnegie Mellon University Tyler Smith, Adventium Labs William Nichols, Carnegie Mellon University ***Critical Technologies: How is the DoD Protecting These Valuable Assets?*** Erin Butkowski, US Government Accountability Office

Rear Admiral Kurt Rothenhaus, USN—is a native of New York City, New York. He received his commission in 1992 upon graduating from the University of South Carolina where he earned a Bachelor of Science degree. He also earned a Master of Science in Computer Science and a Ph.D. in Software Engineering from the Naval Postgraduate School and transferred into the Engineering Duty Officer community in 2003.

His operational assignments include serving as the combat systems/C5I officer on USS Harry S. Truman (CVN 75) and chief engineer on USS O'Brien (DD 975). Additionally, he served on the staff of Destroyer Squadron 15 and on USS Fife (DD 991). He completed an Individual Augmentee tour in Baghdad, Iraq.

His shore tours include: program manager for PMW 160, the Navy's Tactical Networks Program Office, at PEO C4I and commanding officer of Space and Naval Warfare Systems Center Pacific. He also served as the deputy program manager for the Navy Communications and GPS Program Office (PMW/A 170), the assistant program manager for the Consolidated Afloat Network Enterprise Services (CANES) in PMW 160, and the Maritime Tactical Command and Control (MTC2) assistant program manager in the Navy Command and Control Program Office (PMW 150).

His personal awards include the Legion of Merit, Meritorious Service Medal, Joint and Navy and various unit and service awards.

Defensive Industrial Policy: Cybersecurity Interventions to Reduce Intellectual Property Theft

Chad Dacus—is a professor of cyber warfare studies with the Air Force Cyber College. He also serves as the Director, Air University Cyber Research Task Force. His research interests include economics and its contribution to national security and the protection of intellectual property. His writing has been published in journals such as IEEE Security & Privacy, Defense Acquisition Research Journal, and Strategic Studies Quarterly. He holds a PhD in economics from Rice University and an MS in statistics from Texas A&M University. [chad.dacus@au.af.edu]

Carl (Cj) Horn—is a professor of cyber warfare studies with the Air Force Cyber College. Previously, Cj served as the Director of the School of Joint Strategic Studies within the College of Information and Cyberspace at Fort McNair. In that position he led the nation's only War College program centered on information and cyberspace warfare. He served more than 24 years in the U.S. Army as an armor officer and strategist. Cj earned a BS in history from the U.S. Military Academy at West Point and his MA and PhD in history from The Ohio State University. [carl.horn.4@au.af.edu]

Abstract

Through cyber-enabled industrial espionage, China has appropriated what Keith Alexander, the former Director of the National Security Agency, dubbed "the largest transfer of wealth in history." Although China disavows intellectual property (IP) theft by its citizens and has set self-sustained research and development as an important goal, it is unrealistic to believe IP theft will slow down meaningfully without changing China's decision calculus. China and the United States have twice agreed, in principle, to respect one another's IP rights. However, these agreements have lacked any real enforcement mechanism, so the United States must do more to ensure its IP is better protected from China's sophisticated hackers. We call for selective interventions in nascent industries—especially those with important implications for national defense. U.S. policymakers must consider both the supply and demand aspects of the "market" for intellectual property theft to make informed decisions as to how to steer resources. This paper offers insight that the supply side of the equation has been given relatively short shrift. We offer a spectrum of potential interventions to address underinvestment in cybersecurity leading to IP theft and discuss where to go from here.

China's miraculous growth over the past 50 years has lifted hundreds of millions out of poverty. However, some of this growth has taken place at the expense of U.S. corporations who have fallen victim to intellectual property theft on an unprecedented level (Jamali & O'Connor, 2020). Until 2018, the U.S. response has primarily consisted of threats to impose sanctions and indictments of Chinese nationals who are not subject to U.S. jurisdiction. Even in 2018, when the United States imposed tariffs on Chinese goods, the Trump administration's rationale was based on unfair trade practices related to the forced transfer of U.S. technology and intellectual property rather than in retaliation for IP theft (U.S. Trade Representative, 2018). Although an economic and trade agreement signed in 2020 promises some progress on the issue, it should not be presumed that IP theft will slow down significantly (U.S.-China, 2020). Recent reports of increased criminal arrests for IP theft within China certainly provide some room for hope (The National Law Review, 2021). However, a recent policy change prevents the singling out of China because the initiative was being used as a catch-all for cases involving China and led to accusations of bias toward Asian Americans and Chinese citizens (Leslie & Liu, 2022). Given that the United States must continue under the assumption that IP theft will continue, the question becomes whether additional initiatives are necessary to help stem the flow of ideas out of the country.

This paper proposes that more be done to support U.S. organizations' cybersecurity efforts. That is, the focus of U.S. policy should turn to the supply side of the IP theft equation.

Toward that end, we begin with a brief description of the damage to the U.S. economy caused by Chinese IP theft and then proceed to outline both the economic phenomena that can cause private organizations to underinvest in cybersecurity and the benefits the Chinese accrue from pilfering IP. In this way, we lay the groundwork for the cost-benefit framework that follows. This conceptual relationship between the costs of stealing IP versus the benefits of having access to it serves as the initial inspiration for investigating the "supply side" (or U.S. innovation side) as the primary direction for policy change. This thought process is then reinforced by numerous, largely unsuccessful attempts by U.S. policymakers to address China's behavior through demand-side interventions. Finally, we present a spectrum of potential policy innovations designed to address the issue by strengthening the cybersecurity of domestic innovators. Throughout the paper, we alternate our focus between examination of domestic and Chinese phenomena and motivations, and this approach will ultimately lead to the conclusion that more strenuous efforts should be undertaken on the domestic front.

Cost to the U.S. Economy

Estimating the cost of China's cyber-enabled industrial espionage to the U.S. economy is a difficult exercise for a myriad of reasons—not the least of which is lack of specific data. For this analysis, estimates of order of magnitude will suffice. In 2019, the U.S. Patent and Trademark Office estimated that economic activity in IP-intensive industries contributed 41% of gross domestic product. In addition, the same report stated that about 44% of U.S. jobs in 2019 were in industries either directly or indirectly supported by IP-intensive industries and that these jobs paid an average of 60% higher salaries than those in non-IP intensive industries (Toole et al., 2021). The 2013 IP Commission estimated the annual cost to the U.S. economy to be comparable to the current volume (at that time) of exports to Asia, $300 billion (The Commission on the Theft, 2013). Meanwhile, the 2017 IP Commission Report cites a lower bound of $225 billion and an upper bound of $540 billion (The Commission on the Theft, 2017). Converting these figures to 2022 dollars, the inflation-adjusted bounds are $259–$621 billion. To provide an idea of scale, the revenue of the entire U.S. software market is estimated to be $314 billion in 2022 (Statista, 2022). Numerous experts, including Paul Goldstein, have sensibly cast doubt on the accuracy of these estimates, but even using a more conservative estimate backs General Alexander's assessment of an unprecedented transfer of wealth (Goldstein, 2018).

The economy-wide impact of IP theft should not completely overshadow the effects on nascent individual industries. To take one example, Chinese IP theft relating to solar panels was primarily responsible for the bankruptcy of nearly 30 U.S. manufacturing firms. To add insult to injury, many of these firms received government support through subsidies and tax incentives intended for nascent firms involved in the development and provision of energy that is less harmful to the environment ("Made in China," 2019). Another study found that the U.S.-China trade war resulting, in part, from rampant IP theft has likely contributed to a 25.7% increase in bankruptcies in the U.S. farm industry (Wu & Turvey, 2020). According to a study by the Ponemon Institute, nearly 85% of the value of the Standard and Poor's 500 is represented by intangible assets (which include both IP rights and reputation; Ponemon Institute, 2020). In particular, small businesses cannot withstand losses of this magnitude even if the losses are not sustained as cash outlays—attracting investors becomes an impossible task. The Bureau of Labor Statistics has reported that close to 50% of businesses fail within five years, so the margin of error is quite small (U.S. Bureau of Labor Statistics, 2022). Regardless of the reliability of loss estimates, the industries involved and the sheer volume of activity in innovation-intensive industries in the United States should illustrate the importance of addressing this problem more effectively.

Developing Supply and Demand for IP Theft

Supply Side: Underinvestment in Cybersecurity for Information Goods

Not only is malicious cyber activity ubiquitous, but multiple phenomena lead the private sector to underinvest in cybersecurity, thus aggravating the situation. Analysts often cite the frequent presence of externalities in the market for cybersecurity as a theoretical rationale for underinvestment. The negative externality, in this case, is the incurring of losses by those who were not responsible for securing the information technology assets that were compromised by the malicious actor(s). To take just one example, this can happen when a computer is infected and becomes a part of a botnet that victimizes thousands or perhaps even millions of other computers. In addition to externalities, several other related and unrelated theoretical explanations exist for underinvestment in cybersecurity, particularly concerning information goods, as Hal Varian and others have referred to them (Varian et al., 2004; Nabipay, 2018).

Information goods often involve low marginal costs, technological lock-in, network bundling of applications, and network effects. Each of these phenomena can foster market power. First, marginal costs are almost negligible for information goods such as software applications of various types. Another download or search engine query does not cost much to provide. Rising marginal costs have served as a competition-based limit on firm size for generations now, but this competition-enhancing mechanism is often unavailable for information goods. Next, technological lock-in occurs because users are often reluctant to switch platforms once they have adapted to new technology. A famous example of technological lock-in is the baffling long-term dominance of the QWERTY typewriter or keyboard despite its inefficiency. In the modern information age, computer operating systems are arguably the most vivid example. Switching would cost large organizations a staggering sum, and the pain would certainly be felt at the individual level. Once a vendor has you as a customer, the firm is more likely than not to keep you as a loyal customer.

When they apply, network effects can serve as a powerful force for establishing and reinforcing market power. Network effects occur when an application's value depends crucially on the number of other people who use it. Meta (formerly Facebook) is a compelling example of this phenomenon because the entire point of social media is to share ideas and experiences with others. A sparsely populated platform seems almost useless for this purpose. Of course, the more people who use a platform, the more market power that application enjoys. Finally, bundling can help kick-start network effects. For example, Microsoft arguably drove Netscape out of business by bundling its Internet Explorer web browser with its popular operating system. Netscape sued and secured a settlement but lost the war and was acquired by America Online (AOL) in November 1998. AOL stopped supporting Netscape in 2008. As one can see, powerful forces tend to influence markets for information goods to substantial market power or even monopoly. This has important implications for investment in cybersecurity.

With the above conditions often favoring the first (or among the first) application to market, firms face intense pressure to develop the application quickly. Since secure coding coursework is seldom required in computer science programs, these skills are not resident in most application developers' toolboxes without slowing their coding significantly (Lam et al., 2022). Cybersecurity suffers because a secure product that comes to market too late is not likely to garner much market share due to existing network effects and technological lock-in, so the obvious motivation is to rush it to market and get it secured later. In addition, consumers do not have tools with which to meaningfully assess cybersecurity, though Consumer Reports' Digital Lab is a step in the right direction (Consumer Reports, 2022). Concerns about underinvestment in cybersecurity are especially pronounced for smaller, less capitalized firms, with an estimated 43% of cyberattacks directed toward them (Steinberg, 2019). As mentioned earlier, these same firms may not survive a data breach that calls the exclusivity of their intellectual property into question. Now that we have discussed the phenomena that lead to less-than-optimal cybersecurity for potential U.S. victims of IP theft, let us now examine the benefits that accrue to the Chinese when they are successful in stealing secrets.

Demand Side: Benefits of IP Theft to China and Structural Impediments to Reform Within China

Part 2 of China's latest five-year plan focuses on innovation-driven development (Center for Security and Emerging Technology, 2020). IP theft can help jump-start these efforts. The question becomes whether China is well-positioned to take advantage of the acquisition of new knowledge.[49] Several observations (not necessarily a comprehensive list) point to the Chinese faring quite well as consumers of IP theft: (1) a highly educated workforce with particular strength in product development, (2) research and development (R&D) expenditures among the world leaders and unsurpassed expenditure during the experimental development phase of R&D, (3) a socialist economy that can facilitate the transfer of the purloined secrets to those who can use it most efficiently, (4) a Chinese monetary policy that has not, until arguably recently, been focused on a strong renminbi, and (5) China's extensive experience in technology transfer and IP theft.

Among nations not defined as high-income economies, China's workforce has no peer. China stands at an impressive 12th place in the Global Innovation Index, which measures factors as wide-ranging as human capital and research, business sophistication, and infrastructure (Dutta et al., 2021). According to the Center for Security and Emerging Technology (Zwetsloot et al., 2021), China produced 46% more PhD engineering graduates than the United States in 2019 and is expected to nearly double the U.S.'s number of graduates by 2025. Although there is certainly doubt as to the relative quality of doctoral graduates in the two countries, this level of production of scholars and advanced practitioners is impressive. On this measure, China is unambiguously well-positioned to take full advantage of innovations conceived in the United States.

Although all economic data from China should be treated with a fair degree of skepticism, the Organisation for Economic Cooperation and Development (OECD) reports that China spends more on R&D than any other country except for the United States. While the United States leads the world in basic and applied research expenditure, China tops world expenditure on the last stage of R&D, experimental development. Furthermore, China's experimental development spending comprises more than a staggering 82% of its overall R&D investment (Organisation for Economic Cooperation and Development, 2021). From these data, China is arguably the best-positioned country in the world to take advantage of innovations produced elsewhere.

Deng Xiaoping popularized the phrase "socialism with Chinese characteristics" (Moak & Lee, 2015). While the Chinese approach incorporates elements of market economics, its core economic system follows the general tenets of socialism. China's five-year plans are modeled after the Soviet economic model and provide a much more detailed blueprint for economic planning than anything produced by governments in countries with market-based economies. Indeed, China has the world's largest number of state-owned enterprises (Wang, 2021). One can safely assume since the Chinese government supports these enterprises, it can set R&D

[49] A disclaimer on this discussion is that IP thieves lack the understanding involved in making the discovery themselves. The difficulties posed by this lack of knowledge are complex and beyond the scope of this research effort.

priorities within the enterprises it owns to maximize the application of any purloined IP. Furthermore, it is difficult to imagine that any Chinese enterprise could resist significant pressure from its authoritarian central government. We conclude that China's government can steer stolen IP to those it deems best positioned to use it efficiently.

China pegs the value of the renminbi to the U.S. dollar. By definition, countries peg their currencies to an anchor currency to stabilize the exchange rate and minimize exchange rate risk. However, China's motivation to do this is far weaker than it is for smaller economies with less stable economies because it has one of the largest economies in the world and a well-established currency. China has always denied manipulating its currency to keep it artificially low against the U.S. dollar. However, it is undeniable that China's economy has been buoyed by its world-leading level of exports, and a weak renminbi serves to lower the price of its exports. China's most visible actions in setting the exchange rate have served to drop the value of its currency to its lowest point in years (Feng, 2019). At a minimum, China's pegged exchange rate provides it with the opportunity to put its products in the best competitive position possible for garnering market share.

Technology transfer has long been identified as a potential accelerant of economic growth for developing countries (Gurbiel, 2002). Chinese companies are armed with a variety of methods to facilitate the transfer of advanced technologies, including foreign direct investment and joint ventures with foreign companies, venture capital investments, licensing agreements, and talent acquisition. The Chinese government often directs the acquisition to take place and actively assists (O'Connor, 2019). To provide a rough idea of scale, the Chinese participated in 10–16% of all venture deals from 2015 to 2017 (Brown & Singh, 2018). With a population of more than 1.4 billion, the lure of China's large market often proves irresistible to U.S. firms, leading to forced technology transfer. With more than 35 years of executing technology transfer, the assessment that the Chinese have mastered the art of taking technologies and adapting them for their production and use is most likely a gross understatement (U.S. Congress, Office of Technology Assessment, 1987).

China's world-leading number of PhD engineering graduates with unsurpassed experimental development R&D funds to back them is uniquely positioned to take advantage of America's basic and applied research. These well-funded and capable engineers work for organizations that are commonly experienced with technology transfer. In addition, China's powerful central government can steer any IP gains to those who can use it most efficiently and use monetary policy to enhance the cost competitiveness of its products. China denies that it condones IP theft, but this is not the case, based on American indictments of Chinese hackers and the reports of a long-standing commission to address it. China's plausible deniability is further eroded by its unwillingness to cooperate with the prosecution of IP violators. It is difficult to believe that China will slow its IP theft considerably until it becomes more difficult to acquire or there is little left of value that the Chinese do not already possess.

Cost-Benefit Visualization for Intellectual Property Theft

IP theft involves the theft of unique innovations, so an analysis based on a common static analysis of supply and demand is immediately problematic. However, this does not prevent investigating the supply and demand sides of the market in terms of cost-benefit analysis from the point-of-view of the potential thief. Benefit represents the value of the IP to the adversary that is attempting to exfiltrate it. Exactly how this might be measured is an interesting question. Projected future cash flows resulting from the acquisition are the most straightforward and intuitively reasonable metric to use, but uncertainty is likely to be pronounced. If the central government is using the hacker(s) as an agent in a particular case, the hacker may be the one estimating the perceived value to their government employer in a particular instance.

Government officials’ objectives could focus on cash flows but could also consider the prestige involved with acquiring the breakthrough. Regardless, estimated future cash flows generated from the innovation will serve as a useful proxy even if it is not the exact metric used for measuring benefit. Uncertainty is assumed to be significant. Meanwhile, cost represents the difficulty involved in acquiring the IP. The effectiveness of the cybersecurity measures the firm has in place can serve as a very useful proxy for cost. Cost is measured at the “portfolio” level, as the hacker(s) may acquire information about one or more innovations while traversing the firm’s network(s). Uncertainty will also be significant here but is likely to be smaller than the uncertainty involved with estimating benefit. As more information is gained through reconnaissance, this uncertainty may decrease markedly.

Figure 1 depicts a notional relationship between estimated cost and benefit for individual or firm-wide IP theft from China’s point of view. The points represent either individual advances or portfolios of innovations depending on what can be exfiltrated essentially simultaneously (i.e., from a particular firm). For ease of exposition, the scales of the two axes are assumed to be identical. If the estimated benefit of the new technology is greater than or equal to the estimated cost to acquire it, then the clear choice is to attempt to eventually exfiltrate the data. For those observations near the break-even line, the calculus gets a bit murkier because of the high degree of uncertainty involved for both estimates. Since the uncertainty of benefit may well be larger than the uncertainty as to the cost, potential hackers are likely to refrain from attempting to acquire the information when the cost and benefits are nearly equal. Of course, a philosophical point is raised by all of this: What if the Chinese are attempting to hack essentially every U.S. company of any renown with multiple hackers? This supposition assumes that the Chinese have an essentially limitless number of hackers, which seems unlikely but cannot be dismissed out of hand. We will grant that it is likely that many large corporations are likely to be active targets of Chinese hackers, but it is far from clear that China’s hacking labor pool is inexhaustible. Therefore, we assume the Chinese face this dilemma especially for opting to hack specific small businesses, and that the additional costs are meaningful to China’s leadership and an actual choice is made based on a cost-benefit calculus.

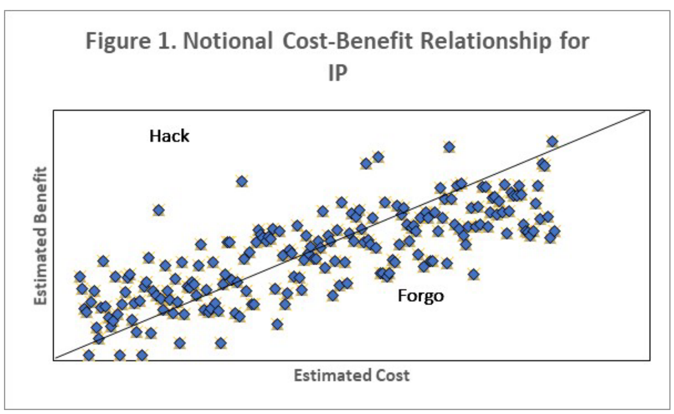


Figure 1. Notional Cost-Benefit Relationship for IP

The goal of policy is to move innovations from the “Hack” area to the “Forgo” area of the chart. To accomplish this movement, either the benefit must go down or costs must increase. Another immediate question is whether we prefer certain innovations to others when considering additional protections through policy. Now that we have a conceptual model for the decision of whether to attempt to exfiltrate trade secrets, we now turn our attention to policy interventions, and we will discover the majority of U.S. policy interventions to specifically combat

IP theft have been designed to address Chinese behavior rather than to increase the cost by making the IP more difficult to pilfer.

U.S. Policy Interventions to Address Intellectual Property Theft

We begin by investigating China's history of weak IP protection and attempts by the United States to change this mindset. The basis for the conclusion of Chinese cultural indifference to intellectual property rights goes back over 2,500 years ago to Confucianism (Alford, 1997). Since Confucian philosophy is inherently collectivist, it is perhaps unsurprising that individual property rights would not be emphasized within China. Moreover, Marron and Steel (2000) identify collectivist values and developing-country status as inversely related to respect for intellectual property rights as reflected in software piracy. The United States itself has a checkered history of protecting intellectual property while it was a developing country (Peng et al., 2017).

In more recent years, the United States has harshly criticized China for its lack of protection of IP rights. This is not a particularly new source of contention between the countries—as far back as 1991 and repeatedly in the 1990s, the United States threatened to impose sanctions on China under Section 301 of the Trade Act of 1974 (Zeng, 2010). With the maturation of the Internet, Chinese IP theft remained a contentious issue, leading to a 2015 agreement between the nations not to "conduct or knowingly support cyber-enabled theft of intellectual property" (White House, Office of the Press Secretary, 2015). After some initial progress, Chinese IP theft continued at its previous pace. The United States has repeatedly indicted individual Chinese hackers for stealing intellectual property (Department of Justice, 2021). In addition, during the Trump administration, the United States levied $200 billion in tariffs on Chinese imports (U.S. Trade Representative, 2018). In 2020, the two countries entered into a Phase One agreement to protect IP, and the United States has already stated that the Chinese are failing to live up to its commitments (Lawder, 2021). In summary, the United States has been assertive in addressing Chinese IP theft from the demand side, with repeated legal, economic, and diplomatic efforts to protect U.S. innovation from the prying eyes of the Chinese. We now turn our attention to the supply side of the equation. What has the United States done to raise the cost involved in stealing IP?

U.S. efforts to specifically protect IP focus primarily on legal remedies and recoveries. The United States has strong legal protections against IP theft against prospective thieves within its borders. The United States is ranked as the leading worldwide protector of IP according to the U.S. Chamber of Commerce's Global Innovation Policy Center (2022). However, IP rights do not extend beyond the borders of the country and rely on effective IP protection laws in the country where it is used. Despite recent efforts to improve IP protection within China, there are strong structural impediments to progress (Rechtshaffen, 2020). Beyond legal remedies, the U.S. government could provide cybersecurity assistance to raise the cost to hackers to exfiltrate IP data.

The U.S. government supports numerous general initiatives to improve the cybersecurity of U.S. firms but little that is specific to IP protection. While examining (or even listing) every cybersecurity initiative of the U.S. government is far beyond the scope of this research effort, some of the more notable activities for our purposes are the design and implementation of cybersecurity standards,[50] cybersecurity threat actor information-sharing programs, free cyber hygiene services, and technical guidance resources provided primarily by the private sector (Department of Homeland Security [DHS], 2022). For critical infrastructure such as the defense industrial base, the Department of Homeland Security (DHS) and sector risk management agencies are primarily charged with ensuring the continuous availability and provision of critical resources and functions through Presidential Policy Directive-21 (PPD-21). Although improved cybersecurity, in general, will certainly also help protect IP, PPD-21 focuses effort exclusively on the security and resilience of critical infrastructure (White House, Office of the Press Secretary, 2013). The authors were unable to find specific cybersecurity initiatives for IP beyond what is offered for general consumption. Notably, the Small Business Administration recently announced a small grant program for bolstering the cybersecurity infrastructure of emerging small businesses (U.S. Small Business Administration, 2022).

The Government Accountability Office's (GAO) recent report—DoD Critical Technologies: Plans for Communicating, Assessing, and Overseeing Protection Efforts Should be Completed—merits discussion on specific protection of IP. The DoD has broken its processes for identifying and protecting critical acquisition programs and technologies into four steps, including identifying, communicating, protecting, and assessing and overseeing the security of critical of technologies. This initiative appears to be a worthwhile extension of the DoD's current role as Sector Risk Management Agency for the defense industrial base. Steps such as including contract language for enhancement of protection efforts can certainly raise the level of protection afforded to these critical technologies (GAO, 2021). The DoD's Protecting Critical Technology Task Force has selected four promising lines of effort including "protecting the research and development enterprise, which includes academia, labs, and universities," but how much progress has been made is unclear (Lopez, 2019). The White House published the first National Strategy for Critical and Emerging Technologies (2020) and has provided an updated list of critical and emerging technologies. The Critical and Emerging Technology Update report states that a strategy on U.S. technological competitiveness and national security is forthcoming. Presumably, this follow-on strategy will contain more definitive prioritization and funding information than this report omits (National Science and Technology Council, 2022). Although these efforts are promising, gaps are likely to remain that can hopefully be partially addressed with the initiatives identified in this paper.

The United States has frequently intervened to influence Chinese behavior regarding IP theft. In addition, U.S. government organizations stand ready to facilitate the legal efforts of aggrieved parties in international courts. However, efforts to strengthen the cybersecurity of innovators specifically to protect IP are lacking. Since the overall effectiveness of U.S. efforts to stem the flow of secrets out of the country has been universally regarded as relatively unsuccessful, we turn our attention to what might be done to improve the situation through cybersecurity assistance.

Protecting Our Most Valuable IP: Defensive Industrial Policy

Since attempting to stem Chinese IP theft through influencing Chinese decision-making seems to have largely failed to this point, the United States should move to shore up defenses. This will involve the shifting of resources to selective nascent industries for defensive purposes.

[50] Note that the primary standards, promulgated by the National Institute of Standards and Technologies, was designed for use by critical infrastructure owners and not necessarily for the consumption of all of private enterprise.

We describe this action as defensive industrial policy, which is distinct from the established concept of industrial policy. Industrial policy is defined as "government intervention in a specific sector which is designed to boost the growth prospects of that sector and to promote the development of the wider economy" (Dadush, 2016). For defensive industrial policy, rather than attempting to boost growth prospects, the purpose is to protect the growth prospects of a nascent sector and the resulting development of the entire economy. The obvious question is how to choose which innovations to protect. We return to our supply and demand framework to answer this question.

To get a good sense of innovation occurring within the United States, one can start with the federal agency responsible for granting patents, the U.S. Patent and Trademark Office (USPTO). Although the patent, copyright, and application for patents and copyrights data this office maintains will represent far from a full picture of U.S. innovation, the inventor's financial interest acts as a powerful incentive to apply for a patent for new technology that can or will be marketable in the near future. However, software developers may not choose to file a patent because the process is too lengthy, costs an average of $50,000, and may not be reliably enforceable against infringement (Chang Villacreses, 2020). In addition, there is no single data field within the patent and copyright data to quickly identify that a particular innovation involves a particular technology, such as artificial technology. Perhaps ironically, Giczy et al. (2022) found the need to use a sophisticated machine learning approach for a recent analysis of artificial intelligence patents. Considering these disclaimers, patents, copyrights, and applications for each could be useful as an initial input into what technologies the United States should prioritize.

For innovations that may have national security implications, the USPTO performs an initial screening and, if national security concerns are evident, refers the application to the appropriate agency. For prioritizing innovations important to national security, this data could be invaluable. In addition, this data set is likely to be much less cumbersome than the full data sets maintained by the USPTO. Of course, this process is unlikely to be free of error, and some inventions may eventually become important or be revealed as important to national security, but this should serve as a basis from which to begin the analysis. As mentioned earlier, the National Science and Technology Council (2022) has generated a list of critical and emerging technologies that could also prove to be invaluable.

Another avenue for identifying important technologies is to use the words and actions of the Chinese. China's five-year plans (FYP) sketch the social and economic development initiatives planned by the Chinese Communist Party and can be a helpful, if perhaps somewhat lagging, indicator of China's R&D priorities. The question becomes whether these plans are predictive of what industries Chinese hackers choose to target. To establish the veracity of this link, we can compare industries identified in China's 12th FYP to Department of Justice indictments from 2014 to 2018 (the alleged thefts took place between 2011 and 2015). We choose to use indictments rather than other sources, such as news articles containing accusations, so that clear attribution rests on a relatively solid foundation. Table 1 lists the industries identified in the FYP with companies named as targets (if specified).

Table 1. 2011–2015 U.S. Industries Allegedly Targeted by Chinese Hackers
(Central Committee of the Communist Party of China, 2011, DoJ, 2014; DoJ, 2017; DoJ, 2018).

Industry Identified in FYP	Year(s) Indictment Occurred	Company(ies)
Energy conservation	2014	SolarWorld
New generation IT	2018	Multiple (unspecified), MSS Cloudhopper
Biological	2018	Unspecified
High-end equipment	2017, 2018	Boeing, Trimble (GPS), Unspecified
New energy	2014	SolarWorld
New material	2014, 2018	Westinghouse, Unnamed
Petrochemical	2018	Unspecified
Light	2014	Unspecified
Textiles	2014	DuPont*
Maritime	2018	Huntington Ingalls
Iron and steel	2014	U.S. Steel
Non-ferrous metals	2014	Alcoa
Building materials	2014	DuPont

These results confirm broad agreement between China's stated policy targets and the illicit activity of Chinese hackers for these years. This data is admittedly dated, and it remains to be seen whether this agreement between policy and hacking behavior will continue. Nevertheless, using the FYPs appears to be a fruitful way to identify IP that may require additional protections. Of course, this should not be the only source of information on China's targets for IP theft. Intelligence reports and investigations of industry claims of Chinese IP theft could also prove quite helpful. If national security concerns are to be prioritized, the industries and technologies identified as ripe targets could be evaluated for their potential importance to national security. Now that we have identified some ways to select industry segments to protect, we turn our attention to what kind of interventions might be helpful.

Spectrum of Interventions to Protect IP

Since the United States has pursued the Chinese on IP theft to a relatively strenuous degree and with underwhelming results, the U.S. government should actively consider policies to strengthen the protection of valuable IP on a technical level. The level of analysis and data required to arrive at the preferred portfolio of policies is beyond the scope of this paper, but we will sketch a general outline of the spectrum of interventions that may prove beneficial. The status quo will represent the lower end of the spectrum with far more active interventions occupying the opposite extreme. Information-sharing efforts and the new DoD process for protecting critical technologies described by the GAO (2021) should be included as a matter of course.

As discussed earlier in the paper, the United States is already actively involved in public-private initiatives to shore up cybersecurity, particularly regarding critical infrastructure. While there is not much in the way of assistance for IP protection, in particular, the critical infrastructure initiatives involve industries likely to have produced and to continue producing the innovations that will protect future national security and fuel economic prosperity. For example, PPD-21 names the Defense Industrial Base Sector as a critical infrastructure sector with the DoD serving as its Sector Risk Management Agency (White House, Office of the Press Secretary, 2013). Under PPD-21, the Secretary of Homeland Security

> evaluates national capabilities, opportunities and challenges in protecting critical infrastructure; analyzes threats to, vulnerabilities of, and potential consequences from all hazards on critical infrastructure; identifies security; identifies security and resilience functions that are necessary for effective public-private engagement with all critical infrastructure sectors; develops a national plan and metrics, in coordination with SSAs and other critical infrastructure partners; integrates and coordinates Federal cross-sector security and resilience activities; identifies and analyzes key interdependencies among critical infrastructure sectors; and reports on the effectiveness of national efforts to strengthen the Nation's security and resilience posture for critical infrastructure. (White House, Office of the Press Secretary, 2013)

This is certainly an impressive list of support activities, but (1) protection of IP is not mentioned as a priority, rather the focus is squarely on resilience (though these security efforts would help with the protection of IP too), (2) the list of duties is so numerous as to be arguably overwhelming and makes it questionable whether the DHS and the Sector Risk Management agencies can truly accomplish all of the duties to more than a superficial level, and (3) the list of identified critical infrastructure sectors is quite lengthy itself, again placing enormous demands on the DHS and sector risk management agencies. Risks that involve the stealing of IP but do not threaten the functionality of critical infrastructure are likely to be discounted due simply to a lack of available personnel and/or funds to address the concern. Aside from critical infrastructure protection, the federal government supports numerous other initiatives that have spillover benefits to the protection of IP, but all of them pale in comparison to its efforts to protect critical infrastructure.

Beyond the status quo, the federal government could provide cybersecurity grants. As mentioned earlier, the U.S. Small Business Administration (SBA) announced it will grant $3 million to strengthen the cybersecurity infrastructure of new small businesses (SBA, 2022). A similar program for smaller businesses with promising IP might help small businesses avoid theft that might put them out of business. In 2018, small businesses accounted for 43.5% of U.S. gross domestic product (Kobe & Schwinn, 2018). Meanwhile, according to The Small Business Guide to Cybersecurity (SCORE, 2020), up to 71% of cyberattacks occur at businesses under 100 employees. There is a lot of variability in these estimates from different sources and years—CPO magazine estimates that 50% of all cyberattacks (Powell, 2019). Regardless, most estimates indicate malicious cyber activity aimed at small businesses outstrips the businesses' contribution to the GDP. Alarming statistics abound when studying the behavior of our adversaries versus small businesses. According to Barracuda Networks (2022), the average employee of a small business with fewer than 100 employees fields 350% more social engineering attempts than an employee of a larger enterprise. Although it is impossible to prove, available data suggests that small businesses underinvest in cybersecurity. Juniper Research (2018) found that small businesses make up only 13% of the overall cybersecurity market. Meanwhile, small businesses make up 99% of all U.S. businesses (SBA, 2020). It seems that adding a grant program specifically for companies with promising IP to the U.S. government's existing efforts could be a step in the right direction. Although Information Sharing and Analysis Organizations (ISAO) are not government organizations, the formation of an ISAO focused on the protection of intellectual property would be beneficial. These organizations can help spread threat-specific information and best practices to innovators across the country.

The U.S. government could go beyond grants to provide some active assistance itself. For example, the U.S. government could field specialized teams to either consult with promising small businesses or provide cybersecurity services to businesses directly. At the extreme, this would be much more interventionist and require a high degree of trust between the parties. This would put these teams into direct "competition" with private-sector cybersecurity service providers but might still prove helpful to businesses that may not be informed consumers of these services.

Many of the same services provided to critical infrastructure providers could also be provided to those organizations with valuable IP. These services include a multitude of basic services provided to the public with many additional services available to critical infrastructure providers (Cybersecurity + Infrastructure Security Agency, 2021, Fall). DHS could survey these services and determine which to promulgate in a more specialized form to organizations with valuable IP.

The final question to address is how any aid might be prioritized between sectors. Reflecting on the government's approach to cybersecurity assistance writ large, critical infrastructure sectors immediately spring to mind. In addition, selecting national security as the paramount national interest for cybersecurity assistance makes sense for several reasons. First, a nation whose security is threatened will be much less able to protect any of the other national interests including ensuring the prosperity of its citizenry. Second, defense spending, in general, has been found to increase economic growth (Sheremirov & Spirovska, 2022), so spending designated to protect R&D gains with national security implications would intuitively boost economic growth to an even greater extent.

Conclusion

China has been stealing U.S. IP for decades. Through examination of both the costs and benefits of IP theft through a supply and demand approach, we have found that the United States has focused almost exclusively on attempting to reduce Chinese hacking through diplomatic and economic means rather than shoring up its own corporations' cybersecurity. We recommend augmenting the current approach to include cybersecurity initiatives aimed specifically at protecting IP. We have provided a spectrum of possible interventions spanning from the status quo to grant programs, additional training materials, and providing specialized teams to actively assist with shoring up cybersecurity for IP protection.

References

Alford, W. (1997). *To steal a book is an elegant offense: Intellectual property law in Chinese Civilization.* Stanford University Press.

Barracuda Networks. (2022). *Spear phishing: Top threats and trends.* https://assets.barracuda.com/assets/docs/dms/Spear-phishing-vol7.pdf

Brown, M., & Singh, P. (2018). *China's technology transfer strategy: How Chinese investments in emerging technology enable a strategic competitor to access the crown jewels of U.S. innovation.* Defense Innovation Unit Experimental.

Bureau of Labor Statistics. (2022, March 7). Business employment dynamics: Establishment age and survival data. Total private. Bureau of Labor Statistics. https://www.bls.gov/bdm/us_age_naics_00_table7.txt

Center for Security and Emerging Technology. (2020). *Proposal of the Central Committee of the Chinese Communist Party on drawing up the 14th five-year plan for national economic and social development and long-range objectives for 2030* (English translation). Georgetown University. https://cset.georgetown.edu/wp-content/uploads/t0237_5th_Plenum_Proposal_EN-1.pdf

Central Committee of the Chinese Communist Party. (2011, January). The 12th five-year plan for economic and social development of the People's Republic of China (English translation). British Chamber of Commerce in China.

Chang Villacreses, D. (2020, November 4). *Protecting your software ideas: to copyright or patent.* https://otc.duke.edu/news/protecting-your-software-ideas-to-copyright-or-to-patent/

Consumer Reports. (2022, February 22). Consumer Reports joins non-profit cyber coalition [Press release]. https://www.consumerreports.org/media-room/press-releases/2022/02/consumer-reports-joins-non-profit-cyber-coalition/#:~:text=New%2C%20global%20organization%20of%20nonprofits,help%20people%20protect%20their%20data&text=YONKERS%2C%20N.Y.,nonprofits%20dedica

Cybersecurity + Infrastructure Security Agency. (Fall 2021). *Services.* https://www.cisa.gov/sites/default/files/publications/FINAL_PDFEPUB_CISA%20Services%20Catalog%202.0.pdf

Dadush, U. (2016, February 1). Industrial policy: A guide for the perplexed. Carnegie Endowment for International Peace. https://carnegieendowment.org/2016/02/01/industrial-policy-guide-for-perplexed-pub-62660

Department of Homeland Security. (2022, March 30). *Shields up*. Cybersecurity and Infrastructure Security Agency. https://www.cisa.gov/shields-up

DoJ, Office of Public Affairs. (2014, May 19). U.S. charges five Chinese military hackers for cyber espionage against U.S. corporations and a labor organization for commercial advantage [Press release]. https://www.justice.gov/opa/pr/us-charges-five-chinese-military-hackers-cyber-espionage-against-us-corporations-and-labor

DoJ, Office of Public Affairs. (2017, November 27). U.S. charges three Chinese hackers who work at internet security firm for hacking three corporations for commercial advantage [Press release]. https://www.justice.gov/opa/pr/us-charges-three-chinese-hackers-who-work-internet-security-firm-hacking-three-corporations

DoJ, Office of Public Affairs. (2018, December 20). Two Chinese hackers associated with the Ministry of State Security charged with global computer intrusion campaigns targeting intellectual property and confidential business information [Press release]. https://www.justice.gov/opa/pr/two-chinese-hackers-associated-ministry-state-security-charged-global-computer-intrusion

DoJ. (2021, November 19). Information about the Department of Justice's China initiative and a compilation of China-related prosecutions since 2018. https://www.justice.gov/archives/nsd/information-about-department-justice-s-china-initiative-and-compilation-china-related

Department of the Treasury. (2019, August 5). Treasury designates China as a currency manipulator [Press release]. https://home.treasury.gov/news/press-releases/sm751

Dutta, S., Lanvin, B., Leon, L. R., & Wunsch-Vincent, S. (2021). *Global innovation index 2021.* World Intellectual Property Organization. https://www.globalinnovationindex.org/userfiles/file/reportpdf/gii-full-report-2021.pdf

Feng, E. (2019, August 5). China's currency falls to lowest exchange rate in 11 years. *NPR*. https://www.npr.org/2019/08/05/748155575/chinas-currency-falls-to-lowest-exchange-rate-in-11-years

GAO. (2021). *DoD critical technologies: Plans for communicating, assessing, and overseeing protection efforts should be completed.* https://www.gao.gov/assets/gao-21-158.pdf

Giczy, A., Pairolero, N., & Toole, A. (2022). Identifying artificial intelligence (AI) innovation: A novel AI patent dataset. *The Journal of Technology Transfer, 47*, 476–505.

Goldstein, P. (2018, April 10). Intellectual property and China: Is China stealing American IP? [Interview by S. Driscoll.] https://law.stanford.edu/2018/04/10/intellectual-property-china-china-stealing-american-ip/

Gurbiel, R. (2002). *Impact of innovation and technology transfer on economic growth: The Central and Eastern Europe experience.* Warsaw School of Economics.

Jamali N., & O'Connor, T. (2020). "U.S., China's cold war is raging in cyberspace, where intellectual property is a costly front." *Newsweek*. September 16, 2020.

Juniper Research. (2018, August 8). Cybersecurity breaches to result in over 146 billion records being stolen by 2023. https://www.juniperresearch.com/press/cybersecurity-breaches-to-result-in-over-146-bn

Kobe, K., & Schwinn, R. (2018). *Small Business GDP: 1998–2014.* U.S. Small Business Administration. https://cdn.advocacy.sba.gov/wp-content/uploads/2018/12/21060437/Small-Business-GDP-1998-2014.pdf

Lam, J., Fang, E., Almansoori, M., Chatterjee, R., & Raj, A. G. (2022). Identifying gaps in the secure programming knowledge and skills of students. *SIGGSE 2022*, pp. 1–7.

Lawder, D. (2021, April 30). U.S. says China has fallen short on "Phase 1" intellectual property commitments. *Reuters*. https://www.reuters.com/business/us-says-china-has-fallen-short-phase-1-intellectual-property-commitments-2021-04-30/
Leslie, R., & Liu, B. (2022, March 4). U.S. Justice Department puts an end to controversial China Initiative. *Lawfare*. https://www.lawfareblog.com/us-justice-department-puts-end-controversial-china-initiative
Lopez, C. T. (2019, November 26). "Task Force Curbs technology theft to keep Joint Force strong." *DoD News*. https://www.defense.gov/News/News-Stories/Article/Article/2027555/task-force-curbs-technology-theft-to-keep-joint-force-strong/
Made in China 2025 and the future of American industry: Hearing before the Senate Small Business and Entrepreneurship Committee. (2019, February 29). 116th Cong., 1st Session. https://www.sbc.senate.gov/public/_cache/files/0/9/090fe492-3ed9-4a1a-b6c1-ebdecec39858/1AB7520770B9032F388CC9E94C79321B.glaser-testimony.pdf
Marron, D., & Steel, D. (2000, April). Which countries protect intellectual property? The case of software piracy. *Economic Inquiry, 38*(2), 159–174.
Moak, K., & Lee, M. (2015). *China's economic risk and its global impact.* Palgrave Macmillan.
Nabipay, P. (2018). *Essays on economics of information goods* [Doctoral dissertation, University of Minnesota]. http://conservancy.umn.edu/handle/11299/202166.
National Science and Technology Council. (2022, February). *Critical and emerging technologies list update*. Fast Track Action Subcommittee on Critical and Emerging Technologies. https://www.whitehouse.gov/wp-content/uploads/2022/02/02-2022-Critical-and-Emerging-Technologies-List-Update.pdf
O'Connor, S. (2019). *How Chinese companies facilitate technology transfer from the United States.* United States-China Economic and Security Review Commission. https://www.uscc.gov/sites/default/files/Research/How%20Chinese%20Companies%20Facilitate%20Tech%20Transfer%20from%20the%20US.pdf
Organisation for Economic Cooperation and Development. (2021). *Gross domestic spending on R&D.* Organisation for Economic Cooperation and Development [Data set]. https://data.oecd.org/rd/gross-domestic-spending-on-r-d.htm
Peng, M., Ahlstrom, D., Carraher, S., & Shi, W. (2017, March). History and the debate over intellectual property. *Management and Organization Review, 13*(1), 15–38.
Ponemon Institute. (2020). *Financial impact of intellectual property & cyber assets: 2020 Aon-Ponemon global report.* Aon-Ponemon Institute. https://www.aon.com/getmedia/6e200c08-c579-4333-b5f2-385ab6fbefde/Financial-Impact-of-Intellectual-Property-
Powell, M. (2019, June 25). 11 eye-opening cyber security statistics for 2019. *CPO Magazine*. https://www.cpomagazine.com/tech/11-eye-opening-cyber-security-statistics-for-2019/
Rechtschaffen, D. (2020, November 11). How China's legal system enables intellectual property theft. *The Diplomat*. https://thediplomat.com/2020/11/how-chinas-legal-system-enables-intellectual-property-theft/
SCORE. (2020). *The small business guide to cybersecurity.* https://www.score.org/resource/small-business-guide-cybersecurity
Sheremirov, V., & Spriovska, S. (2022). Fiscal multipliers in advanced and developing countries: Evidence from military spending. *Journal of Public Economics, 208*.
Statista. (2022). *Technology markets: Software. United States*. Statista. https://www.statista.com/outlook/tmo/software/united-states
Steinberg, S. (2019, October 13). Cyberattacks now cost companies $200,000 on average, putting many out of business. *CNBC*. https://www.consumerreports.org/media-room/press-releases/2022/02/consumer-reports-joins-non-profit-cyber-coalition/#:~:text=New%2C%20global%20organization%20of%20nonprofits,help%20people%20protect%20their%20data&text=YONKERS%2C%20N.Y.,nonprofits%20dedica
The Commission on the Theft of American Intellectual Property. (2013). *The IP Commission Report: The report of the commission on the theft of american intellectual property.* U.S. Patent and Trademark Office. https://www.nbr.org/wp-content/uploads/pdfs/publications/IP_Commission_Report_Update.pdf
The Commission on the Theft of American Intellectual Property. (2017). *Update to the IP Commission Report: The theft of american intellectual property: Reassessments of the challenge and United*

States policy. U.S. Patent and Trademark Office. https://www.nbr.org/wp-content/uploads/pdfs/publications/IP_Commission_Report_Update.pdf
The National Law Review. (2021, July 27). Criminal arrests for intellectual property crimes in China up 99% in H1 2021. *The National Law Review*. https://www.natlawreview.com/article/criminal-arrests-intellectual-property-crimes-china-99-h1-2021
Toole, A., Miler, R., & Rada, N. (2021). *Intellectual property and the U.S. Economy.* U.S. Patent and Trademark Office. https://www.uspto.gov/sites/default/files/documents/uspto-ip-us-economy-third-edition.pdf
U.S. Chamber of Commerce Global Innovation Policy Center. (2022). *2022 International IP index: Compete for tomorrow* (10th ed.). https://www.valueingenuity.com/wp-content/uploads/2022/02/GIPC_IPIndex2022_Report_Full_v2.pdf
U.S.-China. (2020, January 15). Economic and trade agreement between the government of the United States and the government of the People's Republic of China. https://ustr.gov/sites/default/files/files/agreements/phase%20one%20agreement/Economic_And_Trade_Agreement_Between_The_United_States_And_China_Text.pdf
U.S. Congress, Office of Technology Assessment. (1987). *Technology transfer to China.*U.S. Government Printing Office.
U.S. Small Business Administration. (2020). *2020 Small Business Profile.* https://cdn.advocacy.sba.gov/wp-content/uploads/2020/06/04144224/2020-Small-Business-Economic-Profile-US.pdf
U.S. Small Business Administration. (2022, January 21). SBA Administrator Guzman announces new pilot program to bolster cybersecurity infrastructure of emerging small businesses.
U.S. Trade Representative. (2018, June 15). USTR issues tariffs on Chinese products in response to unfair trade practices. https://ustr.gov/about-us/policy-offices/press-office/press-releases/2018/june/ustr-issues-tariffs-chinese-products
Varian, H., Farrell, J., & Shapiro, C. (2004). *The Economics of Information Technology: An Introduction.* Cambridge University Press.
Wang, T. (2021, June 14). Explainer: Why does China have so many state-owned enterprises? *Chinese Government Television Network (CGTN)*.
White House, Office of the Press Secretary. (2013). Presidential policy directive—Critical infrastructure security and resilience/PPD-21. https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/presidential-policy-directive-critical-infrastructure-security-and-resil
White House, Office of the Press Secretary. (2015, September 24–25). President Xi Jinping's state visit to the United States [Fact sheet]. https://obamawhitehouse.archives.gov/the-press-office/2015/09/25/fact-sheet-president-xi-jinpings-State-visit-united-States
Wu, Y., & Turvey, C. (2020). The impact of the China-USA trade war on U.S. Chapter 12 farm bankruptcies. *Agricultural Finance Review, 81*(3), 386–414. https://doi.org/10.1108/AFR-05-2020-0076
Zeng, K. (2010). *Trade threats, trade wars: Bargaining, retaliation, & American coercive diplomacy.* University of Michigan Press.
Zwetsloot, R., Corrigan, J., Weinstein, E., Peterson, D., Gehlhaus, D., & Fedasluk, R. (2021). *China is fast outpacing U.S. STEM PhD growth: CSET data brief.* Center for Security and Emerging Technology. https://cset.georgetown.edu/wp-content/uploads/China-is-Fast-Outpacing-U.S.-STEM-PhD-Growth.pdf

Digital Engineering Effectiveness

Alfred R. Schenker—works in the SEI's Software Solutions Division, and has worked there for more than 20 years. He works to improve software acquisition and product development practices throughout the armed services and other organizations. He has actively worked in software process, architecture, model-based systems engineering, and metrics. Before joining the SEI, Schenker spent more than 20 years in industry as an active contributor in all phases of product development. Schenker is also an inventor, and has obtained patents for a pressure switch (used in automotive airbag applications) and for a manufacturing process to seal gas inside a vessel. [ars@sei.cmu.edu]

Tyler Smith—has more than 10 years of experience that covers data modeling, software integration engineering, and user interface design. Smith is leading the Army-funded Future Vertical Lift support effort, for which Adventium Labs is providing Model Based Engineering expertise, and FRIGATE, a NASA-funded effort to generate failure recovery plans from traditional engineering models. Smith has led Adventium's data modeling and open API development efforts and spearheaded Adventium's analysis interoperability strategy, which enables rapid integration of AADL-based analysis tools into a wide range of third-party system engineering environments. Additionally, he oversaw the Capstone Pilot Multi-Organization Model Integration exercise that validated AADL-based model integration, ASoT practices, and CVIT. [tyler.smith@adventiumlabs.com]

William Nichols—is a senior researcher at the SEI with a PhD in Physics. His experience includes data acquisition, sci-entific software development, nuclear design and plant simulation, and software project management. While at the SEI he has coached software project management and published research on using soft-ware metrics for project and program management. [wrn@sei.cmu.edu]

Abstract

The 2018 release of the DoD's Digital Engineering (DE) strategy and the success of applying DE methods in the mechanical and electrical engineering domains motivate application of DE methods in other product development workflows, such as systems and/or software engineering. The expected benefits of this are improved communication and traceability with reduced rework and risk. Organizations have demonstrated advantages of DE methods many times over by using model-based design and analysis methods, such as Finite Element Analysis (FEA) or SPICE (Simulation Program with Integrated Circuit Emphasis), to conduct detailed evaluations earlier in the process (i.e., shifting left). However, other domains such as embedded computing resources for cyber physical systems (CPS) have not yet effectively demonstrated how to incorporate relevant DE methods into their development workflows. Although there is broad support for SysML and there has been significant advancement in specific tools, e.g., MathWorks®, ANSYS®, and Dassault tool offerings, and standards like Modelica and AADL, the DE benefits to CPS engineering have not been broadly realized. In this paper, we will explore why CPS developers have been slow to embrace DE, how DE methods should be tailored to achieve their stakeholders' goals, and how to measure the effectiveness of DE-enabled workflows.

Introduction

We, as an engineering community, are designing, assembling, and deploying the most ambitious and complex systems ever made. The details of these systems stretch beyond the ability of one, ten, or even one hundred individuals to comprehend; therefore, we must engineer these complex systems with teams of thousands. Success in pursuits such as these requires systems management—and a key tenet of systems management is measuring progress.

Problem Statement

The emergence of *digital engineering* (DE) has the potential to improve project outcomes (e.g., reduction of acquisition risk for cost and schedule) for cyber-physical systems (CPS) by enabling defect detection to "shift left." *Shifting left* is enabled by developing new methods (e.g., model-based analysis) that discover significant defects earlier in the product life cycle. Allowing defects to escape and not be discovered until final integration and test, means that much more effort will be spent fixing them. Coupling the late discovery of defects with high expectations and pressure from management is the perfect environment for making careless mistakes. It will also likely result in other inefficiencies, reducing the project team's effectiveness, and potentially burning out the workforce. The benefits of DE have been clearly demonstrated in other domains (e.g., nuclear power system design). However, recent studies highlight the challenges of both implementing DE and measuring the DE process for CPSs. Shifting defect detection to the left suggests that requirements and design issues can be found earlier (before test of the code or embedded system). However, there is no guarantee that applying DE methods early in the development life cycle for CPSs and software will result in the improved likelihood of attaining stakeholders' goals. In this paper, we review research on DE for CPSs and recommend how to measure DE methods.

Key Takeaways

Effective application of digital engineering is challenging. If we approach digital engineering as a box to check, we will fail. Instead, engineering leaders should:

- Establish individual and organizational goals that are qualitative, rather than quantitative.
- Establish measurement objectives early and foster a culture that encourages rigorous process.
- Incorporate tight feedback loops between the digital engineering artifacts and the evolving system design. Use the models to reduce system development risks.
- Clearly differentiate what you track from what you measure.
- Do not attempt to measure digital engineering application to cyber-physical systems development *purely* by counting defects. Instead, employ proven systems engineering design and testing practices, and use digital engineering to accelerate and improve them, rather than replace them.

Background

CPSs—also called embedded computing systems—are "engineered systems that are built from, and depend upon, the seamless integration of computational algorithms and physical components" (National Science Foundation, 2021). In the last three decades the category of CPS has grown to include virtually all automobiles and aircraft.

Research Context

Although the capabilities of modern systems are unparalleled, the costs—particularly the costs of software—are rising at an unsustainable rate. The amount of software used in modern aircraft is growing exponentially (as shown in Figure 1), and the cost of software is steeply rising as well. This situation leads to the question, "How can we avoid exponential growth in cost if software size is growing exponentially?"

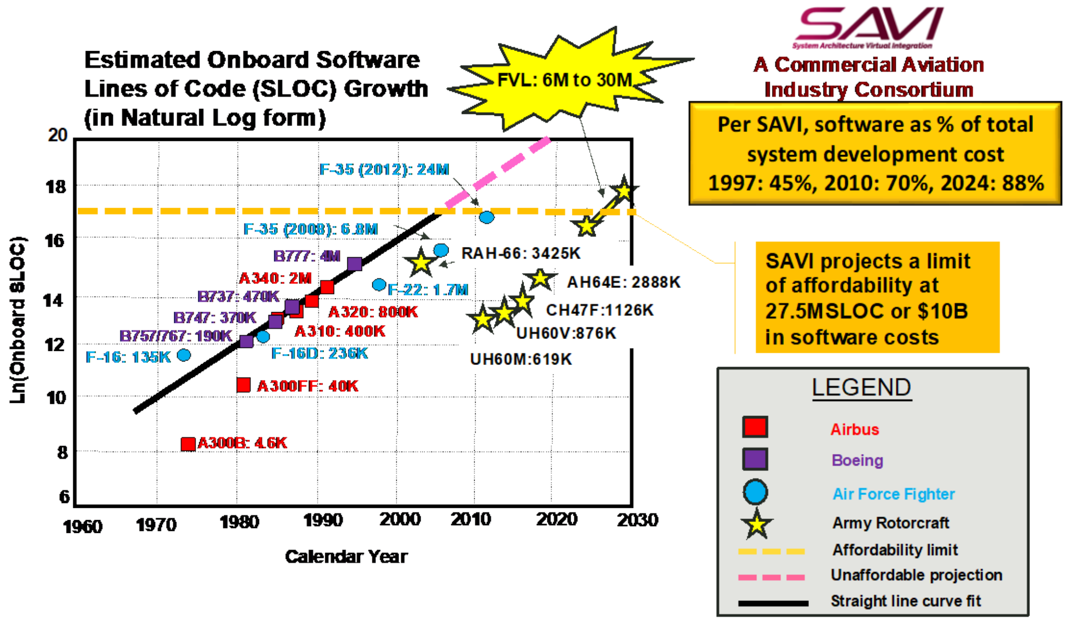


Figure 3. The Growth of Software Lines of Code in Aviation (Lewis, 2019)

At the beginning of the space race in the late 1940s, the Army, Air Force, and their contractors grappled with the challenges of the unprecedented complexity of large-scale rocketry. Early efforts (e.g., the Jet Propulsion Laboratory [JPL] *Corporal* rocket, which was developed and tested in the late 1940s and 1950s) were frequently stymied by integration errors and component failures. The ad hoc approach to building and testing components that had worked for simpler systems was no longer sufficient (Johnson, 2002, p. 86). Just like today, engineers and engineering managers were faced with exponential increases in complexity and insufficient tools to manage that complexity.

JPL addressed this problem by adding systems engineering rigor (i.e., design reviews, change control, configuration management), which contributed to the success of NASA's Mariner program in the early 1960s and 1970s (Johnson, 2002, p. 108).

Today's challenges in embedded computing system complexity echo the early challenges of the space race. Once again, we face a major leap in system complexity and we must adapt or face major cost overruns and missed deadlines, just as experienced by the Joint Strike Fighter program (GAO, 2021). The Mariner program demonstrated that systems engineering rigor could help programs contend with complexity, and the emerging practice of DE has the potential to meet our complexity needs now.

History of Digital Engineering

The Department of Defense (DoD) defines Digital Engineering (DE) as "an integrated digital approach that uses authoritative sources of system data and models as a continuum across disciplines to support life cycle activities from concept through disposal" (DoD DES, 2018). DE has been with us—in some form or another—since the 1960s. The Apollo program relied heavily on simulations for both astronauts and mission control personnel to reduce risk to operations (Kranz, 2000). Modern manufacturing is driven by computer-aided design (CAD) tools that allow high-fidelity design and evaluation before bending metal. Modern aircraft like the Boeing 787 are extensively modeled and evaluated using digital tools to evaluate everything from flight dynamics to manufacturing. There is no doubt that DE works for *some* domains of engineering. For others, particularly CPSs, the answer is less clear.

There was a revolution in automated manufacturing when suppliers of factory automation equipment began to adopt 3-D (CAD) tools. The change occurred gradually, starting with Sales and Marketing; imagine the impact on a proposal when yours is the only one showing off a 3-D model of a new assembly line. The change then slowly became integrated into the machine design process. Eventually, it was possible to conduct design reviews using modeling tools, and animations could be constructed to show the sequence of operations. These changes enabled many different stakeholders (e.g., machine builders, machine operators, maintenance mechanics, safety engineers, etc.) to "see" the virtual machine months before it was built. Through the design review process, many stakeholders could modify the design to suit their unique needs, in ways that would have been much more difficult using legacy methods.

Applying Finite Element Analysis (FEA) to the mechanical design process illustrates how a legacy method can be used to verify and validate a design, even as the design process undergoes a metamorphosis. FEA was first commercially developed in the late 1960s with help from NASA as an open-source product called NASTRAN. As the mechanical design process has evolved—from paper to 2-D CAD to 3-D CAD—analysis tools have evolved as well. In many cases these analysis tools can be integrated into the design process, providing the opportunity for a tight feedback loop to experiment with applying mechanical loads to the surfaces in order to observe the virtual effects, with the design updated as a result of the analysis.

The motivation to incorporate model-based analysis methods into the design process was influenced by the consequences of not applying them, e.g., late discovery of design flaws or architectural issues such as the Ariane 5 Launch failure (Dowson, 1997) or the Therac-25 radiation therapy machine failure (Leveson & Turner, 1993). The lead time associated with redesigning after late discovery could result in a serious impact on the program schedule; often these changes involved tooling with months of lead-time. Clearly, the means to perform virtual prototyping was necessary as a risk mitigation mechanism.

When defects are discovered late, the first impulse is to try to fix it with software. There is a common perception that software has no lead-time. Although software can be written quickly, depending on the complexity of the CPS, the implementation of the redesign might have unintended consequences. Ultimately a fix that was first thought to be resolvable through only software might require changes to other parts of the system. For example, a change to a data element format (as measured by a CPS) might require changes to the device driver, but also may impact the GUI and/or the system diagnostics. So, a "simple" fix might take weeks to implement and, in response to schedule pressure, might be implemented sub-optimally.

Nuclear plant design provides another recent example of successful DE use. Modern engineering tools simulate the physics of reactor designs. Simulation tools separately design and analyze structural mechanics, neutron transport, coolant flow, and heat transfer. The simulations are used to test and iterate on design decisions in a feedback loop, without the need to construct physical prototypes. Using overall plant simulation, nuclear plant designers can test operating conditions and lifetime performance. Construction has also benefited from 3-D CAD that enables the fabrication and assembly of complex systems in limited and confined spaces. In all cases, models can be used as the medium for evaluating changes or for external review, enabling or accelerating feedback loops in the design process.

Keys to the success of existing DE approaches include not only physically faithful simulations, but also interfaces that enable virtual and incremental design with an Authoritative Source of Truth (ASoT). In nuclear plant design, physical modeling propagates the hydraulic channels and material properties to neutron transport and fluid flow. Physical structure and neutron flux propagate to heat analysis. Structure, fluid flow, and flux propagate into plant analysis. Each model reuses the design from other models but makes local simplifications or includes model-specific additions. For example, heat transfer analysis can lead to changes in the nuclear fuel structure. Each characteristic has an ASoT, and each change can be analyzed as part of a design process feedback loop. Operational conditions can be tested, leading to redesign that begins with fuel properties and distributions. The tools available have a profound influence on an effective design workflow.

The Promise of Digital Engineering for CPSs

The promise of DE for CPSs, as with many other domains of systems engineering, is to "shift left." As shown in Figure 2, defects are much less expensive to fix when they are found on the left side of the traditional engineering "V" model. This benefit has been demonstrated conclusively for physical systems.

DE of CPSs should enable similar feedback loops to those observed in domains like nuclear power system design or FEA. An initial design includes components with interfaces, properties, and constraints. A virtual integration (analogous to the simulated behavior of a nuclear power plant) could either verify that the components are compatible or identify necessary design changes. A virtual analysis would simulate the execution of the virtual system to verify that (1) the constraints were satisfied and (2) inputs produced the expected outputs. Failure in this virtual test would require recycling of the design. It is important to note that this virtual integration and analysis does not replace more traditional methods of modeling and simulation, it provides the means to perform the tests—and find the issues—earlier.

The economic case for DE based on early defect detection was argued by Feiler et al. (2013) and Hansson et al. (2018). In short, these researchers present evidence that, in the domain of embedded safety-critical systems, 35% of errors are introduced in requirements, and 35% of errors are injected in architectural design. Nonetheless, 80% of all errors are not *discovered* until system integration or later.

According to conventional wisdom, the cost of correcting an issue in later phases can be 300 to 1000 times the cost of correcting it in earlier phases (Dabney, 2003). As the systems grow, the escalating cost of late discovery overwhelms development costs. (See Figure 2).

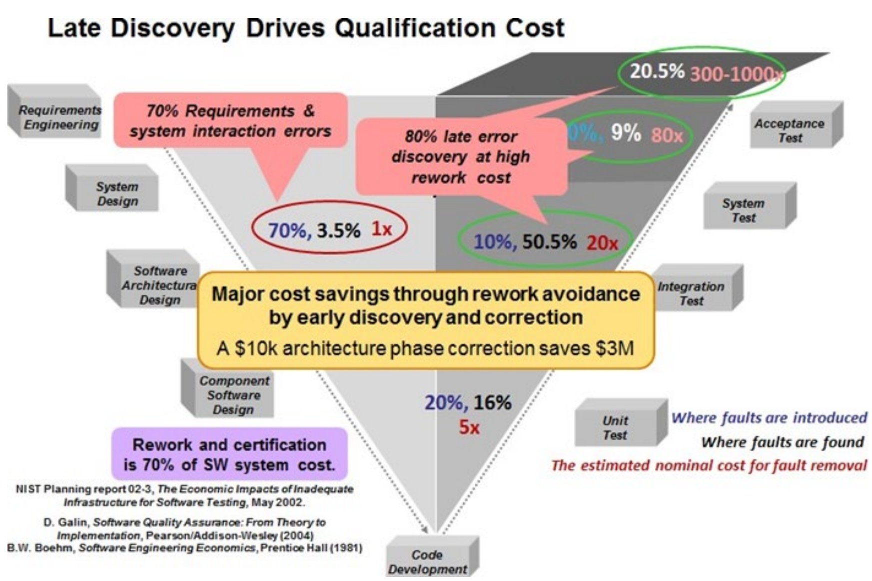


Figure 4. Gap Between Defect Origin and Discovery (Feiler et al., 2013)

Recent Department of Defense Digital Engineering Initiatives

DoD Initiatives

In recent years, Congress mandated the DoD adopt a Modular Open Systems Approach (MOSA) for systems development, which directed procurement officials to pursue modularity in CPSs to reduce costs across families of systems.[51] The Air Force is pursuing an Agile software strategy to reduce software integration costs across platforms (Roper, 2020, p. 15). The Army spearheaded development of Architecture Centric Virtual Integration Process (ACVIP) to find cyber-physical integration errors early through virtual integration (Boydston et al., 2019). In FY2020 Congress also mandated that the DoD "establish a digital engineering (DE) capability to be used (A) for the development and deployment of digital engineering models for use in the defense acquisition process; and (B) to provide testing infrastructure and software to support automated approaches for testing, evaluation, and deployment throughout the defense acquisition process."[52]

The DoD is planning significant investment in DE, including providing a framework in the 2018 *Digital Engineering Strategy* (DES) (DoD DES, 2018). The DES relates five expected benefits from DE:

1. informed decision making/greater insight through increased transparency
2. enhanced communication
3. increased understanding for greater flexibility/adaptability in design
4. increased confidence that the capability will perform as expected
5. increased efficiency in engineering and acquisition practices

The *Digital Engineering Strategy* calls for practitioners to, "[e]stablish accountability to measure, foster, demonstrate, and improve tangible results across programs and [the] enterprise" (DoD DES, 2018). Recent efforts such as the Comprehensive Architecture Strategy (CAS) provide a framework for measurement by formalizing the relationships between key business drivers, key architecture drivers, and quality attributes for a CPS (CCDC, 2018). Now, a new way is needed to measure those quality attributes, as called for in the DES. Without methods to measure DE effectiveness, it will be difficult for new programs like the Army's Future Vertical Left (FVL) to gauge whether they are on track to reap the benefits of DE. Effective measures should provide systems managers with the information they need to determine how well they are meeting each of these five benefits.

U.S. Army JMR MSAD

The SEI and Adventium Labs participated in the Army's Joint-Multi-Role Mission System Architecture Demonstration (JMR MSAD) Science and Technology (S&T) program, which ran from 2013–2020. This program exercised and evaluated DE tools and standards for CPSs, particularly tools oriented toward ACVIP. The program included three demonstrations, all conducted with significant industry participation and contributions:

1. The Joint Common Architecture (JCA) Demonstration (2014–2015) demonstrated the use of the Future Airborne Capability Environment (FACE™) Technical Standard and

---

[51] See NDAA 2021 section 804.B.iii
[52] See NDAA 2020 section 231

Joint Common Architecture (JCA) Functional Reference Architecture (Wigginton, 2016).

2. The Architecture Implementation Process Demonstrations (AIPD; 2017–2018) evaluated approaches to model-sharing among organizations.
3. The Capstone demonstration (2018–2020) evaluated a larger portion of the system development lifecycle, including the use of DE tools to adapt a CPS design to new requirements (Jacobs et al., 2021).

The JMR MSAD program highlighted major new capabilities in DE but also brought the problem of measurement to the fore. The government asked performers to evaluate whether new DE tools and approaches were effective; however, the answer was unclear. Most performers tracked the effort (person-hours) required to perform various engineering tasks. However, these measurements did not yield clear results, for several reasons:

- Performers had to learn to use DE tools and had little time to separate learning from using. This made it difficult to compare the effectiveness of the new process against baselines.
- Performers were using multiple new technologies simultaneously, such as the Future Avionics Capability Environment (FACE^{TM}) and Architecture Analysis and Design Language (AADL). This confounded the challenge of learning these technologies and made measuring their effectiveness difficult as the “learning” period was not well defined.
- Performers integrated a mix of new and existing cyber-physical components but did not clearly differentiate the engineering costs and benefits of applying DE between new and existing components. We expect greater gains from applying DE on a *new* component during its design phase than from applying DE to an existing component after its design is complete.
- The schedule of the Capstone exercise likely motivated some performers to delay application of DE technologies until late in the design process, when the effectiveness of DE was likely diminished.

For example, Raytheon, one of the performers in the JMR MSAD Capstone exercise, presented the chart shown in Figure 3 at the program wrap up meeting. Raytheon measured labor hours spent conducting both DE and traditional engineering and software to complete development and integration tasks directed by the Government. Figure 3 refers to specifically to the ACVIP approach to DE.

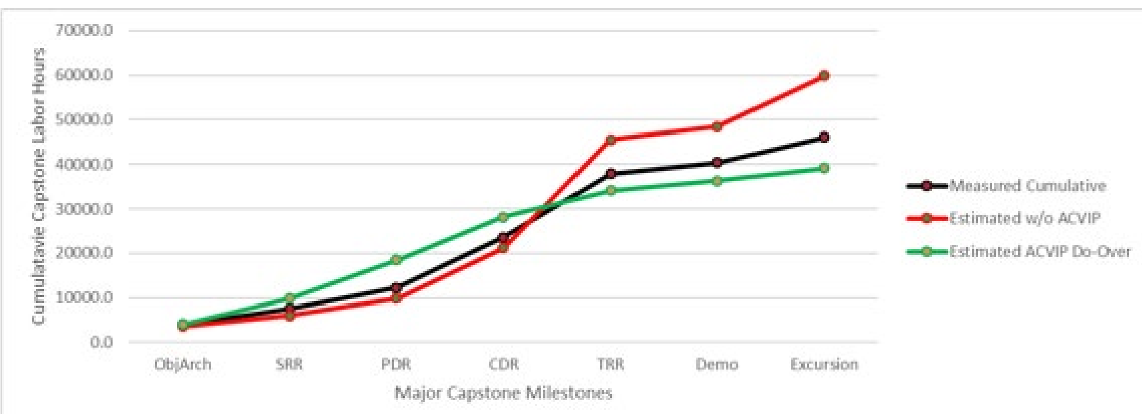


Figure 5 Raytheon, a JMR MSAD Capstone Performer, Provided this Chart as Part of Their Lessons Learned Briefings (Raytheon and General Electric Aviation, 2020)

The structure of the Capstone exercise meant that Raytheon had no definitive baseline against which to measure DE effectiveness, which presumably motivated their use of estimates for the purpose of comparison. These estimates show the expected "shift left," in which there is an up-front cost to DE (around the time of PDR, the red line is lowest) but a larger savings later (at TRR and Excursion the red line is highest). However, the Capstone excursion involved many new DE technologies and (as noted in Figure 3) performers like Raytheon predicted a different effort distribution in a hypothetical repeat of the exercise (green line). Although encouraging for the benefits of DE, these results indicate uncertainty in the digital engineering measurements employed in the JMR MSAD Capstone exercise.

The Capstone exercise performers used a combination of new and existing components as input to the integration exercises. This meant that, in some cases, performers conducted DE after a component design had already been completed, limiting the opportunities for DE to influence the component development process.

The DEVCOM Capstone Final Report noted the need for further work in defining the concept of a "defect" in the context of DE (Jacobs, et al., 2021). As discussed later in the Are We Learning From Our Mistakes section, defect tracking can be an effective measurement approach to evaluate digital engineering, but only with a clear definition of defect. For example, some Capstone Mission System Integrators (MSIs) found that a supplied component did not behave as expected. The component was not necessarily *broken* but its behavior model (used to facilitate DE activities for the component) was not consistent with its actual behavior (Jacobs, et al., 2021). Was this a defect in the implemented component? A defect in the model? The answer is not obvious. The definition of defect used for DE effectiveness measurement needs to be sufficiently precise to avoid ambiguity in such a situation.

Continued Army Investment

Ongoing DoD efforts, such as the Army's Future Vertical Lift (FVL) acquisition programs continue to leverage DE via model-based Government Furnished Information (GFI) for performers and including ACVIP in requirements. The methods we propose in this paper will help government stakeholders assess the effectiveness of these techniques as implemented by their suppliers.

Digital Engineering Effectiveness Challenges

Measurement Has Side-Effects

Measurement comes with a few challenges. The problems begin when determining the subject of the measurement. Before measuring something, we must know the goal of measuring it and the decisions the measurement will inform. Is the goal to become more efficient, produce a higher quality product, or something else? What are the implications of that goal? By "higher quality," do we want a product that is longer lasting, has fewer defects in use, or something else altogether? Further, are we using the measurement to compare, select, predict, or evaluate? Simply determining what to measure and why is only the beginning.

We cannot always measure something directly; instead, we often measure something related to the object of the measurement. For example, customer satisfaction is difficult to measure directly; however, usage trends or customer change requests can be useful indicators of the customer's satisfaction. However, two problems result. The first is accounting for confounding and/or complex relationships, and the second is that using measurements to control the system changes the system.

Confounding is a problem encountered when trying to identify cause-and-effect relationships because there are multiple causes and/or multiple effects. Some causes can be hidden, and sometimes causal factors interact. For example, a sidewalk could be wet from rain or from a lawn sprinkler. Just knowing the sidewalk is wet doesn't tell you why it's wet.

A related problem is *complexity*—when causes can have multiple outcomes. For example, ice cream sales may predict shark attacks because both have the common cause of summer heat. It's best to be cautious and understand whether the object of the measurement is a *cause* or a *result*.

The second problem, often called Goodhart's Law, can be paraphrased as "When a measure becomes a target, it ceases to be a good measure." More precisely, Goodhart stated, "Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes" (Goodhart, 1984). We interpret this statement as an observation that using a measurement from the system to exercise control changes the system's behavior. If the relationship is not trivial, the conditions under which the initial behavior was observed will no longer apply.

Therefore, we must understand the causal relationships of the objects of measure with the subjects of measure, including the environment in which the relationships were observed. This could be problematic for DE, where we aim to make substantial changes in workflows.

Another problem is that the measurements we choose might not translate to different goals. Measurement and analysis rely on assumptions about how a system behaves and the environment where the system operates. Assumptions valid for one analysis may not be suitable for another. Story points, for example, are often used as a proxy for relative effort, and sometimes as a proxy for relative size. This isn't such a big problem because cost and effort are related; however, that correlation can ruin an analysis that combines an assumption of size with an assumption of effort. Productivity is the ratio of size and effort. Defect density is the ratio of defects and size. Combining the derived metrics based on different assumptions can therefore lead to unpredictable results.

Finally, measurement plans can be inflexible. Selecting a measurement plan involves tradeoffs between the value of the information and the cost of acquiring measurements. Changes to the workflow can disrupt this balance. Changes in the information needs can require measurements that had been dropped from the original plan. Changing the measurement plan has costs and can introduces delays, therefore, expect resistance to changing an existing measurement plan. However, reusing the existing measure may not be appropriate.

These are just the conceptual challenges. The practical problems remain: instrumentation, accuracy, cost of measurement, and validation. We are not suggesting that measurement is impractical. Instead, we are saying that we must be cautious about simply reusing measurements devised for other purposes and applying them in the new context of DE for CPSs.

Culture Change is Hard

As we change processes to incorporate the lessons learned from DE experiences, we need to recognize that implementing DE is likely to fundamentally change many aspects of the business model. For example, when we describe an activity as "enabling a shift left," we must recognize that we won't be able to achieve shifting left unless the necessary resources are available earlier in the life cycle. So, to build the capability to shift left, we need the same types of engineers (e.g., test engineers, integration engineers, acceptance testers) to be involved much earlier in the process.

This approach seems to make sense, until you view it from a PM's perspective. The PM fears that there will be extra effort needed at the end of the contract to resolve issues that arise when trying to perform final integration and testing. The PM might assume that early life cycle spending to shift left will jeopardize the resources needed at the end of the life cycle. Since the PM might not believe that shifting left will result in fewer issues in final integration and test, they might resist committing these critical resources early.

Perhaps the PM is right, and maybe that is the point of this paper. There is no guarantee that applying DE effort early in life cycle will result in more efficient integration and testing. For example, we can spend lots of time building models and conducting analyses without producing meaningful results.

However, if DE is done well—the modeling and analysis activity focused on design tradeoffs that we know (either from experience or prior analysis) must be done correctly—the organization should be able to uncover and resolve many of the issues that would not normally be discovered until much later (i.e., during integration and test).

Therefore, implementing DE will result in higher quality and more predictable product development. The organization must be committed to the principles embodied in DE and continue to refine their process as they learn how best to apply their effort. Clearly there will be a stratification of competency in contractors as they evolve their processes to incorporate DE artifacts and practices.

What Information Is Relevant to Digital Engineering?

If you don't establish the abstract concepts you want to measure, it will be difficult to interpret your data. You must establish the questions you want to answer before moving on to concrete measurements.

What information do you need to know to run a successful DE effort? For this paper, we focused on goals one, three, and four from the *Digital Engineering Strategy*:

- Goal 1: Informed decision making/greater insight through increased transparency
- Goal 3: Increased confidence that the capability will perform as expected
- Goal 4: Increased efficiency in engineering and acquisition practices

Information Related to Decision Making/Insight

Consider a parent who is annoyed by their child leaving the front door open and allowing flies to get into the house. The parent, wanting to eliminate the flies in the house, devises a plan to reward the child for killing the flies. However, the opposite happens: The child let more flies into the house because they were being paid a bounty for killing them. When the parent changed the incentive to reward the number of consecutive days without letting a fly into the house, the child's behavior changed.

There is a similar conundrum with respect to shift-left processes. By their very nature, we hope to discover much earlier the issues that would "bite us" at the end of the project. It is easy to create measurements to count defects by life cycle phase, but how do we know that the ones we find are the really bad ones—the showstoppers? We won't know until we get to the end of the project whether all the effort we spent to shift left paid off in smoother integration and qualification. We need ways to identify that our upfront investment in DE was worth it.

Information That Capability Will Perform as Expected

As part of the AIPD exercise in 2017 and 2018, participants created, exchanged, and virtually integrated models with one another. Through this activity, we learned that agreement on a modeling language is necessary but not sufficient to enable efficient communication of DE concerns among stakeholders. Additional information specifying modeling patterns and paradigms is required to facilitate effective DE collaboration between organizations. This discovery motivated the creation of the AADL Annex for the FACE 3.0 Technical Standard and a variety of modeling templates used in later Army efforts (SAE International, 2019).

Information Related to Increased Efficiency in Engineering and Acquisition Practices

Where Does Rework Originate, Where Is It Performed?

In practice, counts of defects depend upon how you categorize them and when you choose to count them. Data suggests the relative number of defects injected during construction is high (Vallespir & Nichols, 2012), however, most of these defects are usually detected and addressed in reviews or unit test without being recorded in the defect tracking system. That is, these defects tend to be found near to their origin and are not the expensive defects that delay projects (Boehm, 1981).

Although rigorous unit test is effective for construction defects, it is far less effective at finding requirements or design errors; these continue to be found during integration or system test. This gap in origin and removal of architecture and design errors is thus an opportunity for DE to speed up the process by identifying errors at or close to their origin.

Are We Learning From Our Mistakes?

A leading indicator of an organization's culture is the way that the organization adapts its processes to account for defects or issues that escape in-phase detection. Sometimes, it is unavoidable (i.e., the issue was so complex that the investment in modeling and analysis to find it would have been ineffective). However, many times a retrospective analysis (e.g., a post-mortem review) of the issue points to a potential change to some aspect of the development/test process that will reduce the likelihood that this same problem will recur. Characterizing the types of defects as they are discovered and conducting some form of triage enables the organization to understand where it needs to focus its attention.

Many organizations that seek to apply modeling and analysis methods to their development process already possess the data they need to focus their DE efforts. These organizations have historical records of prior efforts that should help them understand which types of defects are typical, and which are not. Standard techniques, such as Pareto analysis, can be used to organize the data to illustrate where the pain is, and DE experts can recommend methods that address these pain points.

DE methods (e.g., modeling and analysis) are providing new ways to assess product design and architecture earlier in the lifecycle. So, as we perform root cause analysis, we need to ask this question: How could we have found this if we had been using DE methods? The answers to this question should provide the right type of guidance to the organization.

Repenning and Sterman (2001) describe the challenges of workflow and culture change in their article "Nobody Ever Gets Credit for Fixing Problems That Never Happened." In this article, they use the example of the rush to embrace Total Quality Management (TQM) that many organizations adopted in the 1980s and early 1990s. They note that by the mid-1990s TQM appeared to have run its course, and these same organizations were abandoning it. They were now rushing to search for the next "silver bullet."

However, their research showed that "companies making a serious commitment to the disciplines and methods associated with TQM outperform[ed] their competitors." They go on to point out that business leaders are facing a cultural paradox. In a world where it is becoming increasingly easier to identify and learn about new technologies that could improve their performance, the challenge is how to implement the technology the right way. They point out, "You can't buy a turnkey six-sigma quality program. It must be built from within."

The same applies to how model-based methods, such as those characterized as DE, need to be viewed. Each organization has its own challenges and nuances, and the application of the DE methods must be engineered to fit the organization's culture.

How Do You Measure Digital Engineering?

Project management measures are not necessarily the same as measuring the effectiveness of a paradigm shift, in particular the use of DE to "shift left." A major lesson from the JMR MSAD Capstone demonstration was that performers reused measures intended for project management rather than measurement tailored to evaluate the performance difference resulting from a change in methodology.

Measurement of Decision Making/Insight?

Although some processes are straightforward to measure, effective DE is typically a more subjective than objective measure. We can observe what happens, but not why it happens. We could assess the ways in which the DE has been applied. For example:

- How is the hardware modeled (i.e., is it a black box or a white box model)?
- Which computing resources are modeled (e.g., processors, memory, network bandwidth)?
- How are the models used for feedback control, or to verify and validate the system design?
- How do system engineers use the results of the model-based analyses?

These methods are distinctly not quantitative in nature. That is, we can measure some aspects but won't be able to produce a number to put on a scale and claim effectiveness. However, if we evaluate data in the context (with a demonstrated history) that the methods have proven to be effective on prior projects of similar scope, it is possible that a qualitative assessment, such as what is described above, is what we need.

However, measuring the things we are accustomed to measuring (e.g., effort, schedule, quality) might not be effective either. How do you know the fidelity of the modeling effort based only on the hours spent on modeling? We could measure attributes of the models (e.g., Has the model been reviewed?), but how do you know that the quality of the system is where it needs to be at this point in the life cycle? Do you base the measure of quality on the number of defects found?

To make matters more difficult, many organizations do not even count defects until the system goes through a major review (e.g., SRR, PDR). As the requirements are elaborated, (e.g., into use cases or scenarios) and distributed for others to review, no one is tracking the issues that are found as defects—until after the SRR milestone. At that point, the defects are all considered to be "baselined." How is it possible to understand the quality of the evolving system design under these conditions? More importantly, can we use DE tools (e.g., models and analysis methods) to inform the measure of quality?

Fundamentally, measuring changes to the culture or way of doing business is a multifaceted problem the defies simple or narrow measures. Objectively, we can measure activities and outcomes, but we must take care to measure enough different aspects in a sufficient amount that we don't encourage "hitting some target."

Figure 4 represents a way that an organization can use a DE approach to provide feedback and improved insight about the evolving quality of the system during, for example, requirements elaboration.

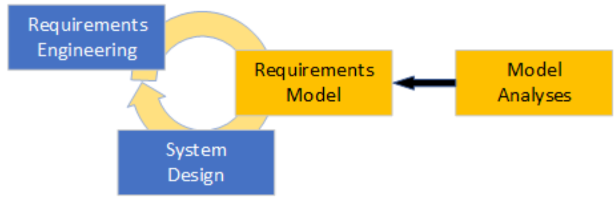


Figure 6. Feedback Control Using DE For Design and Requirements

The use of a requirements model during the requirements engineering life cycle phase provides an opportunity to verify and validate the planned implementation of a requirement in the context of the system, as it evolves. Note that this application is consistent with the traditional, batch-oriented approach typical of weapon system development and the more iterative approaches, such as Agile or DevSecOps. The model provides the representation of the system that informs the system design activity as the requirements are elaborated.

Measurement of Capability Performing as Expected

Following the AIPD exercise, Adventium Labs conducted an Army-funded study on multi-organization DE. Adventium Labs found that comparative tracking of communication channels used between organizations (e.g., phone, email, model exchange) is an applicable method for determining whether your DE methods are sufficient for communication. (A decline in email accompanying introduction of DE tools may indicate that the DE tools are providing sufficient information [Smith et al., 2018].)

When creating DE artifacts of CPSs such as system models, the number of approaches to creating the model can be overwhelming. The number of available modeling styles, tools, and patterns can make it challenging for stakeholders to communicate with one another about what information to provide in DE artifacts. To address this issue, Adventium Labs also developed a collection of “report cards” for evaluating DE applications. These report cards provide a rubric for objectively evaluating DE artifacts. For example, the Model-Readiness Report Card for ACVIP consists of 15 measures that provide indicators about how well a project is applying ACVIP (e.g., Are you using a standardized modeling language? Are you adhering to its semantics?). The report cards can be used for evaluation by a customer, third party, or as a self-assessment.

How Do You Measure Performance?

Identifying the key information need begins with identifying the performance goals. Overarching goals include reducing cost, reducing overall development time, and increasing the resultant predictability of project performance. Although project management measures might be helpful, they are inadequate because they do not account for the multitude of factors that affect cost and schedule. That is, each project is a unique system.

In addition to the indicators for overall cost and schedule, we need more insight into what happens along the way. What were the activities performed, in what order, and what were the results? Where were evaluations performed, and what decisions were made? What are the sources of cost and rework? It’s critical to understand what activities were performed, the ordering of performance, what they produced, and how they contributed to the performance goals of cost and lead time we enabled. Rather than just outputs, a detailed process analysis helps us answer not only what, but how and why.

Analysis of Rework—JMR MSAD

Shift left implies that we should shift effort from test into design and design analysis. This includes not only additional time in design (modeling) activities, but it also includes identifying design defects and implementing design changes. The indicators of successfully shifting left include more work in design, significant effort in design evaluation (virtual integration and virtual analysis), identification and classification of design defects in the design evaluation, and the amount of effort fixing those discovered defects. The key measures, therefore, include the following:

- effort spent on each of the specific activities
- counts of defects found by life cycle phase (e.g., design review, virtual integration, virtual analysis, and later physical integration and test)
- a judgement about which phase the defects were injected in
- the effort required to identify and implement the necessary changes
- categorization of the defects to determine the detection effectiveness of each activity by defect type

On the JMR MSAD Capstone demonstration, our initial attempt at measurement included creating work packages for the requirements, design modeling, virtual integration, and virtual analysis to estimate the effort. Unfortunately, this failed for the following reasons.

First, work was not tracked against specific work packages. Work packages were planned for a specific time period (sprint). At the end of the period, we knew which work was completed and the estimated effort (as a portion of the period effort). However, the actual effort update could only apply a sprint factor to all efforts. That is, the relative effort of the completed work packages to each other were not updated. We had no way to measure a change in relative effort for each work package, thus we could not identify a change in relative effort applied to activities (e.g., design, virtual integration, virtual analysis). This problem could be avoided by tracking work packages individually (e.g., Kanban style) rather than as a batch.

Second, the problem was compounded by failing to account for information needs when changing the workflow and work sequencing. The information need was to isolate effort in the different activities. Traditionally, a single story would progress through the workflow, perhaps noting the time of state change as different activities were completed. Instead, the story was decomposed into separate tasks to isolate the activity effort. However, stories were not tracked individually; they were only tracked as a batch within the sprint. Moreover, the related sub-stories were not executed consecutively and usually not in the same sprint. Thus, not only was all information on actual activity effort lost, but the sum of effort for related stories was also lost. The activity mix of stories changed in each sprint while reported effort was the estimate adjusted by a sprint-based factor. However, we did see that the virtual integration and virtual analysis stories were not occurring at the expected rates. This indicated that the workflow was not conducive to iterative feedback loops. This measurement problem might have been mitigated with a more pipelined workflow in which a design element immediately went into virtual integration or virtual analysis.

Third, we did not track defects. We had no way of knowing how much of the time in virtual integration or virtual design was involved in the time needed to execute the tests versus the time needed to fix the defects found. We could have measured the effort through time spent in the activity, or by specifically identifying a new story as rework resulting from a defect discovered in virtual integration or virtual analysis. There are options for setting up the accounting that interact with the intended workflow. For example, are defects fixed as they are found in the analysis, or are they sent for further evaluation and remediation? The measurement plan requires starting with the measurement goal and then assessing how to measure that goal within the context of the workflow. If the workflow changes, the measurement plan must be revisited.

Measuring Improved Efficiency in Engineering and Acquisition Practices

Properly realized and applied, DE should move defect discovery closer to the defect origin, resulting in fewer issues found in physical integration and test. ACVIP envisions accomplishing this by modeling interfaces and constraints (Boydston et al., 2019). Virtual activities can implement verification activities that identify requirements and design issues before building code or integrating code into the CPS. It is also possible that more rigorously designed products will have fewer defects. In either case we expect fewer design or architectural issues to be discovered during physical integration and test. Also, the additional steps early in the process should shift effort to earlier activities while reducing effort in integration or test.

Before physical construction, the modeled components should be virtually integrated to find architectural defects. In other words, integration issues, if present, should be found during virtual rather than physical integration. A virtual analysis simulates the integration of the components operating as a system to uncover requirements defects. These activities require effort, but they also produce artifacts (issue reports and product changes); fortunately, effort and changes are both measurable and analyzable. Another change in workflow is incremental development with tighter feedback control using virtual integration and analysis.

The virtual integration and analysis should occur whenever components are created or modified, well before software integration, build, or system test. These defects should, therefore, be found more frequently. The reduced time (and scope of other changes) between a defect's injection and discovery in virtual integration or analysis should reduce the cost of identifying the source of the defect and the cost to fix it.

Using ACVIP as part of a DE effort should result in the following:

- incremental model development
- virtual integration and analysis
- more defects discovered in virtual integration and analysis
- fewer defects discovered in physical integration and analysis
- more effort expended prior to integration and test
- less effort expended post integration and test
- a shift in organizational skill sets to support digital tools
- increased exchange of digital artifacts in place of or supporting traditional development artifacts

Table 1. The New Status Quo for the Development of CPSs

<table>
<tr><th>Lifecycle Phase</th><th>Injection %</th><th>Discovery %</th><th>Discovery %</th></tr>
<tr><td>Requirements</td><td rowspan="3">70</td><td rowspan="3">3.5</td><td rowspan="3">80</td></tr>
<tr><td>System Design</td></tr>
<tr><td>Software Architecture</td></tr>
<tr><td>Component Software Design</td><td></td><td></td><td></td></tr>
<tr><td>Code Development</td><td>20</td><td>16</td><td>15</td></tr>
<tr><td>Unit Test</td><td rowspan="2">10</td><td rowspan="2">50.5</td><td rowspan="2">3</td></tr>
<tr><td>Integration Test</td></tr>
<tr><td>Acceptance Test</td><td></td><td>9</td><td>2</td></tr>
<tr><td>Post-Acceptance Test</td><td></td><td>20</td><td>0</td></tr>
</table>

Table 1 represents our view of the new status quo for the development of CPSs. While we don't think it is realistic to eliminate the discovery of defects in integration and acceptance test, we do think that by effectively using DE methods, an order of magnitude reduction in "out-of-phase" defect discovery is possible. Over time, the defect-based effectiveness measurement of DE should show a trend toward the desired defect detection rate shown in Table 1.

At least two model-based engineering mechanisms directly contribute to efficiency: (1) earlier discovery of design issues and (2) shorter feedback loops for learning and correction. A less direct mechanism includes benefits from precise, consistent, and exchangeable design artifacts. The result should be fewer defects, faster fixes of defects, and less total rework.

Figure 5 illustrates our view of how to represent the incorporation of effective DE practice in the context of the traditional system engineering V Model. The outside boxes are consistent with life cycle processes represented in the traditional V Model. The black arrows represent the verification and validation activities that relate the testing to the design. The inside boxes represent DE artifacts in the form of (1) models (or representations) of the CPS or CPS elements and (2) analyses that can be performed on the models.

The analyses are typically developed to answer specific questions about the system's properties (e.g., latency [data], processor utilization, memory utilization). The DE environment provides the system development organization with an early view into the expected behavior of the physical system. With the exception of the Requirements model, the DE artifact, though labeled with different words, represent the same virtual model incorporating higher levels of fidelity. For example, the Software Architecture model might represent the components as black boxes, with just their interfaces modeled; but in the Component model, these black boxes would be replaced with white boxes, and the overall system model would be updated to reflect the higher level of fidelity in that part of the virtual model.

When this approach is coupled with an analysis capability that can be applied recursively (even automatically, if deployed in a CI/CD pipeline), the development organization can identify and react quickly to potential issues, solving the problems in phase. These problems would not normally be found until months (or even years) later, using simulation or hardware-in-the-loop laboratories.

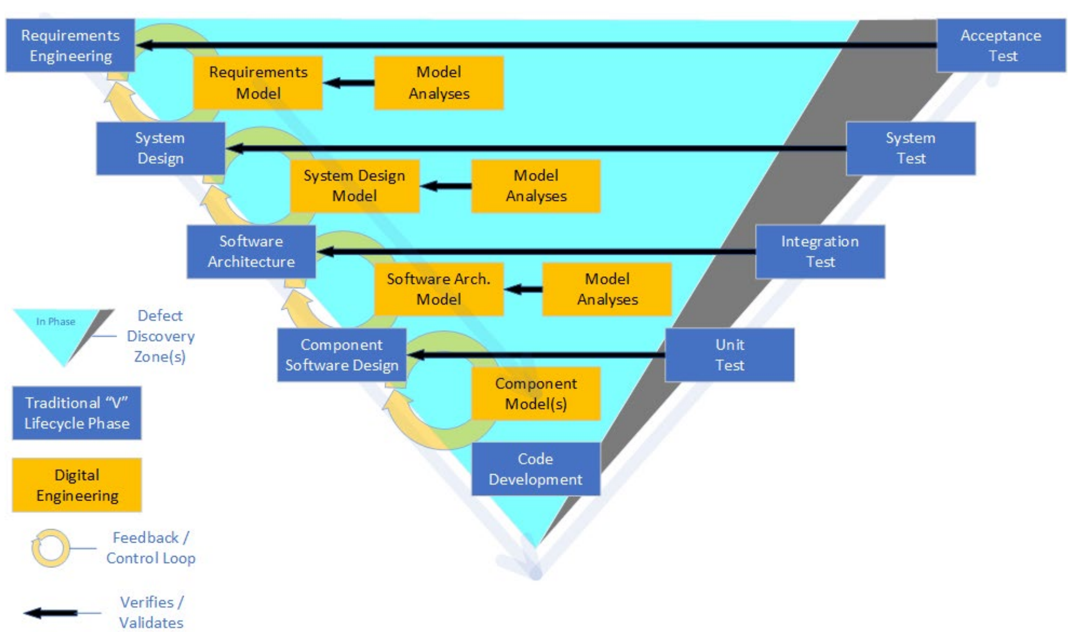


Figure 7. Incorporating Effective DE Practice in the V Model

Finding design issues earlier (preferably in-phase, but at worst in-phase + 1) should reduce total costs. As noted in Section 2.2.1, a substantial number of requirements and design defects are discovered in system test. Historical data in research by Boehm and Menzies shows that test or operational defects have cost factors of 10 to 100 times more than defects found earlier (e.g., in design reviews; Boehm & Basili, 2001; (Menzies et al., 2016). The high cost of late changes is realized in the traditional CPS domain because of the (1) potential for changes to propagate through the system, (2) the difficulty of isolating the root cause of a test failure, (3) the likelihood that personnel who performed the initial work will be unavailable for diagnosing and fixing the problems, and (4) fading memory resulting from the time span between initial development and fixes.

Tighter feedback loops that identify problems close to their creation mitigate all of these problems. Tighter feedback loops also benefit learning and improvement. Ideally, defect analysis includes examining the process to understand why the mistake was made. This is most effective when the original developers are involved, and the analysis is performed promptly. Those most familiar with the development not only are best suited to analyze the root cause of a problem, but also to adjust future practice to avoid similar mistakes. Reducing the time gap between the design and problem evaluation improves the effectiveness of root cause analysis.

Finally, the exchange of artifacts is an opportunity for misunderstanding. The authoritative source of truth reduces errors from redundancy and inconsistency. Digital artifacts can be used directly in reviews rather than tools such as PowerPoint, which require transcription (avoiding the associated transcription errors) and can allow more flexibility and precision when addressing reviewer questions. DE artifacts improve the effectiveness of communicating the design and identifying issues early.

Conclusion

If you take one thing away from this paper, it's that DE is a major change for CPS development, and that measuring the effect and effectiveness of that change requires a methodical approach. Don't focus exclusively on quantitative measures. They can be skewed by myriad factors that are out of direct control. Instead, increase the usage of qualitative measures and be mindful that your measures do not become targets.

What we are trying to get across is how to identify the practices that distinguish an organization that is just "checking the DE boxes" from one that is building the DE culture into its development approach. These distinguishing practices include the following:

- Be mindful of your expectations when measuring. Be ready for up-front costs (i.e., a learning curve) when adopting DE methodology.
- Recognize that culture change will likely be necessary. There are proven methods for overcoming cultural resistance that should be considered as part of the planning for DE.
- Identify separate measures that show whether DE is being used (e.g., tighter feedback loops) from whether it is effective (higher product quality and avoiding cost and schedule overruns).
- Learn how to apply DE practices for maximum effect. Some problems will not be appropriate, others may not warrant the additional effort. Although initially the organization should initially err on the side of over-modeling, the feedback control should indicate whether or not the effort is adding value.

What we learned from Capstone was that much of the DE was applied out of phase (i.e., after the design was completed, and used primarily for verification purposes), so it didn't function as part of a control loop. The measures applied in Capstone were also applied outside of the control loop. The result was measurements with limited confidence.

Opportunities for future research should include evaluating the structured application of DE practices as part of a feedback control loop to inform life cycle design activities. The instrumentation for these evaluations is especially important since the resulting data will likely conclusively show the value of the early application of DE practices.

References

Boehm, B. (1981). *Software engineering economics.* Prentice Hall.

Boehm, B., & Basili, V. R. (2001). Software defect reduction top 10 list. *Computer, 34*(1), 135–137. doi:10.1109/2.962984

Boydston, A., Feiler, P., Vestal, S., & Lewis, B. (2019). Architecture centric virtual integration process (ACVIP): A key component of the DoD digital engineering strategy. *22nd Annual Systems and Mission Engineering Conference.* Retrieved from https://www.adventiumlabs.com/publication/architecture-centric-virtual-integration-process-acvip-key-component-dod-digital

CCDC. (2018). *Comprehensive architecture strategy (CAS) version 4.0.* U.S. Army Combat Capabilities Development Command (CCDC), (Aviation and Missile Center (AvMC). Retrieved March 28, 2022, from https://apps.dtic.mil/sti/pdfs/AD1103295.pdf

Dabney, J. B. (2003). *Return on investment of independent verification and validation study (preliminary phase 2B report).* NASA.

DoD DES. (2018). *Digital engineering strategy (DES)*. Retrieved March 29, 2022, from https://ac.cto.mil/digital_engineering/

Dowson, M. (1997). The Ariane 5 software failure. *ACM SIGSOFT Software Engineering Notes, 22*(2), 84. doi:251880.251992

Feiler, P. H., Goodenough, J. B., Gurfinkel, A., Weinstock, C. B., & Wrage, L. (2013). *Four pillars for improving the quality of safety-critical software-reliant systems.* Software Engineering Institute. Retrieved March 28, 2022, from https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=47791

GAO. (2021). *F-35 joint strike fighter: DoD needs to update modernization schedule and improve data on software development.* Retrieved March 28, 2022, from https://www.gao.gov/assets/gao-21-226.pdf

Goodhart, C. A. (1984). Problems of monetary management: The UK experience. In *Monetary Theory and Practice* (pp. 91–121). Macmillan Education UK. doi:978-1-349-17295-5_4

Hansson, J., Helton, S., & Feiler, P. H. (2018). *ROI analysis of the system architecture virtual integration initiative.* Software Engineering Institute. doi:10.1184/R1/12363080.v1

Jacobs, W., Boydston, A., Ba, P., Rhamy, K., Davis, J., Padilla, M., & Roberson, H. (2021). *Joint multi-role mission system architecture demonstration—Capstone demonstration, government final report.* U.S. Army CCDC Aviation and Missile Center (AvMC).

Johnson, S. B. (2002). *The secret of Apollo systems management in American and European space programs.* The John Hopkins University Press. Retrieved March 28, 2022, from https://www.press.jhu.edu/books/title/2845/secret-apollo

Kranz, G. (2000). *Failure is not an option.* Simon & Schuster. Retrieved March 28, 2022, from https://www.simonandschuster.com/books/Failure-Is-Not-an-Option/Gene-Kranz/9780743214476

Leveson, N. G., & Turner, C. S. (1993). An investigation of the Therac-25 accidents. *Computer. IEEE Computer Society*, 18–41.

Lewis, W. (2019). Tool expo keynote Address—AADL addressing software related integration issues and costs. *Tool Expo for Model-Based Embedded Systems Development.* Huntsville, AL.

Menzies, T., Nichols, W., Shull, F., & Layman, L. (2016). Are delayed issues harder to resolve? Revisiting cost-to-fix of defects throughout the lifecycle. *Empirical Software Engineering, 22*(4), 1903–1935. doi:https://doi.org/10.1007/s10664-016-9469

National Science Foundation. (2021, January 7). *Cyber-physical systems (CPS)*. Retrieved from https://beta.nsf.gov/funding/opportunities/cyber-physical-systems-cps

NIST. (2002). *The economic impacts of inadequate infrastructure for software testing.* Retrieved March 28, 2022, from https://www.nist.gov/document/report02-3pdf

Raytheon and General Electric Aviation. (2020, December 14). *MSAD capstone program wrap-up MSI 3 assault*.

Reppening, N., & Sterman, J. (2001). Nobody ever gets credit for fixing problems that never happened. *California Management Review, 43*(4).

Roper, W. (2020). *There is no spoon: The new digital acquisition reality.* U.S. Air Force. Retrieved March 28, 2022, from https://software.af.mil/wp-content/uploads/2020/10/There-Is-No-Spoon-Digital-Acquisition-7-Oct-2020-digital-version.pdf

SAE International. (2019). *Architecture analysis & design language (AADL) annex F: AADL annex for the FACE™ technical standard edition 3.0.* doi:https://doi.org/10.4271/AS5506/4

Smith, T., Whillock, R., Edman, R., Lewis, B., & Vestal, S. (2018). *Lessons learned in inter-organization virtual integration.* SAE International. doi:https://doi.org/10.4271/2018-01-1944

Vallespir, D., & Nichols, W. (2012). An analysis of code defect injection and pemoval in PSP. *Proceedings of the TSP Symposium 2012*, 3–20, Pittsburgh.

Wigginton, S. (2016). *Joint common architecture demonstration (JCA demo) final report.* U.S. Army RDECOM. Retrieved March 28, 2022, from https://apps.dtic.mil/sti/pdfs/AD1012511.pdf

Critical Technologies: How is the DoD Protecting These Valuable Assets?

Erin Butkowski—is a Senior Analyst with the U.S. Government Accountability Office (GAO). She is a member of the GAO's Contracting and National Security Acquisitions team and has written reports on a variety of topics, including Coast Guard cutters, disaster response contracting, critical technologies, and the DoD's pursuit of artificial intelligence. [butkowskie@gao.gov]

Abstract

Critical technologies—such as elements of artificial intelligence and biotechnology—are those necessary to maintain U.S. technological superiority. As such, they are frequently the target of theft, espionage, and illegal export by adversaries. Prior Department of Defense (DoD) efforts to identify these technologies were considered by some military officials to be too broad to adequately guide protection. This presentation examines (1) the DoD's recent efforts to identify and protect its critical technologies and (2) opportunities for these efforts to inform government-wide protection efforts.

Background

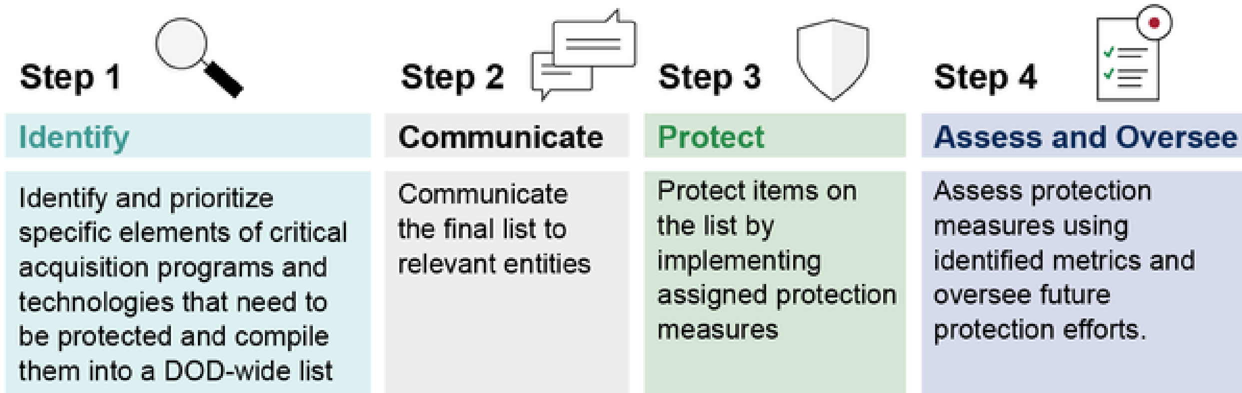
The federal government spends billions annually to develop and acquire advanced technologies. It permits the sale and transfer of some of these technologies to allies to promote U.S. national security, foreign policy, and economic interests. However, the technologies can be targets for adversaries. The John S. McCain National Defense Authorization Act for Fiscal Year 2019 requires the Secretary of Defense to develop and maintain a list of acquisition programs, technologies, manufacturing capabilities, and research areas that are critical for preserving U.S. national security advantages. Ensuring effective protection of critical technologies has been included on the Government Accountability Office's (GAO's) high-risk list since 2007.

Objectives, Scope, and Methodology

This report examines (1) the Department of Defense's (DoD's) efforts to identify and protect its critical technologies and (2) opportunities for these efforts to inform government protection activities. The GAO analyzed DoD critical acquisition program and technologies documentation and held interviews with senior officials at the DoD and other federal agencies responsible for protecting critical technologies.

Summary

Critical technologies—such as elements of artificial intelligence and biotechnology—are those necessary to maintain U.S. technological superiority. As such, they are frequently the target of theft, espionage, and illegal export by adversaries. The DoD has outlined a revised process to better identify and protect its critical technologies, including those associated with acquisition programs throughout their life cycle or those early in development. Prior DoD efforts to identify these technologies were considered by some military officials to be too broad to adequately guide protection. The revised process is expected to address this by offering more specificity about what elements of an acquisition program or technology need to be protected and the protection measures the DoD is expected to implement. It is also expected to support the DoD's annual input to the National Strategy for Critical and Emerging Technologies, which was first published in October 2020.



Source: GAO depiction of Department of Defense's (DOD) process. | GAO-21-158

Figure 1. Overview of the DoD's Revised Process to Identify and Protect Critical Acquisition Programs and Technologies

The DoD began implementing this process in February 2020 with an initial focus on identifying critical acquisition programs and technologies that need to be protected and how they should be protected. As of February 2022, it has not yet determined which metrics it will use to assess the sufficiency of protection measures.

Once completed, the revised process should also inform the DoD and other federal agencies' protection efforts. Military officials stated they could use the list of critical acquisition programs and technologies to better direct resources. Officials from the Departments of State, Commerce, and the Treasury stated that they could use the list, if it is effectively communicated, to better understand what is important to the DoD to help ensure protection through their respective programs.

For additional information, see GAO-21-158 as well as prior related GAO work, including GAO-15-288 and GAO-21-119SP.

PANEL 24. MODERNIZING AND ASSESSING ACQUISITION BUSINESS PROCESSES

Thursday, May 12, 2022	
1:55 p.m. – 3:10 p.m.	**Chair: Lieutenant General Thomas H. Todd III, USA,** Deputy Commanding General, Acquisition and Systems Management ***Leading Practices: Agency Acquisition Policies Could Better Implement Key Product Development Principles*** Shelby Oakley, US Government Accountability Office ***Open for Business: Business Models for Innovation with Modular Open Systems Approaches*** Gregory Sanders, Center for Strategic and International Studies ***Measuring the Effects of Federal Budget Dysfunction: Impacts of Continuing Resolutions on Public Procurement*** Spencer Brien, Naval Postgraduate School Korey W. Letterle, United States Marine Corps. Paul A. Kantner, United States Marine Corps.

Lieutenant General Thomas H. Todd III, USA— began serving as the Deputy Commanding General for Acquisition and Systems and the Chief Innovation Officer at U.S. Army Futures Command in July 2020.

He previously served at the executive level in numerous roles: Special Assistant for Acquisition and Systems Management to the Commanding General, U.S. Army Materiel Command, Program Executive Officer for Army Aviation, Deputy Commanding General of Research, Development and Engineering Command, Senior Commander of Natick Soldier Systems Center, and Modernization Advisor to the Director, Army Capabilities Integration Center.

LTG Todd led Army and Joint programs at all echelons, delivering advanced capabilities to Soldiers, joint services, other government agencies, and over 60 foreign allied nations. He developed and delivered advanced capabilities CH-47F, H-60M, H-60V, AH-64E, Improved Turbine Engine, MQ-1C Gray Eagle ER, UH-72 Lakota, EMARSS-E, and the Black Hawk Aircrew Trainer. His joint assignments include: Defense Contract Management Agency Special Programs Multi-Service Team, Chief of Contracts, Joint Task Force Bravo, Honduras. Operationally, he served with A Co 3/501st Aviation Regiment in the Republic of Korea and 4th Squadron, 6th Cavalry Brigade, Fort Hood, TX.

LTG Todd is a 1989 graduate of The Citadel. He is an honor graduate of the Army's Initial Entry Rotary Wing training and a graduate of the Kiowa and Black Hawk Maintenance Test Pilot courses. He holds Masters of Science degrees in Contract Management and Strategic Studies from the Florida Institute of Technology and the U.S. Air War College, respectively.

His awards and badges include the Legion of Merit (2 Oak Leaf Cluster), the Defense Meritorious Service Medal, the Meritorious Service Medal, the Joint Service Commendation Medal and other Army and joint commendations and awards. He is an Air Assault graduate and a Senior Army Aviator, rated in the UH-1 Iroquois, OH-58 A/C Kiowa, UH-60 A/L/M Black Hawk and CH-47 D/F Chinook.

LTG Todd resides in Austin, Texas with his wife. They have three children and one grandchild.

Leading Practices: Agency Acquisition Policies Could Better Implement Key Product Development Principles

Shelby S. Oakley—is a Director at the U.S. Government Accountability Office, Contracting and National Security Acquisitions. [OakleyS@gao.gov]

Abstract

Each year, the Department of Defense (DoD), the Department of Homeland Security (DHS), and the National Aeronautics and Space Administration (NASA) together invest billions of dollars to acquire complex, hardware- and software-centric systems to provide critical defense, security, and space capabilities. Given the amount of federal funds spent and the critical missions these agencies support, Congress and agencies have consistently underscored the importance of achieving efficiencies and effectiveness across these acquisition activities. The GAO has also contributed to these efforts, and agencies and Congress have acted on many of the GAO's recommendations, including taking steps toward implementing knowledge-based acquisition frameworks, which the GAO's prior work found is essential to improving performance. Nonetheless, the GAO's annual assessments of major acquisition programs at each agency continue to find that programs often take significantly longer, cost more than initially estimated, and in some cases deliver final products with less capability than anticipated. Leading companies would not be able to sustain such outcomes without potentially going out of business. This dynamic correspondingly drives leading companies to undertake a disciplined approach to product development—one that is instructive to government acquisition, despite environmental differences. Throughout an individual product's development, leading companies often confront difficult tradeoff decisions, such as options about design requirements, technical solutions, and where and when to launch a promised solution. These decisions are largely informed by the incentive to be first to market within a globalized marketplace and win enduring customer support.

Why This Matters

Each year, the Departments of Defense (DoD) and Homeland Security (DHS) and the National Aeronautics and Space Administration (NASA) together invest hundreds of billions of dollars to buy stealth jets, cutters and ships, and lunar rovers, among other things, all with complex software. However, the Government Accountability Office's (GAO's) annual reviews of these agencies' major acquisitions find they often take longer and spend more money than planned to deliver capabilities to users.

Key Takeaways

Leading companies take a disciplined approach to develop innovative products that satisfy their customers' needs and to deliver them to market on time and within planned costs. The 13 leading companies the GAO interviewed perform similar activities when developing new products, such as iterative design in hardware and software development. These activities in the development process align with the four key principles that help project teams deliver innovative products to market quickly and efficiently (see Figure 1). The GAO found that the department-wide acquisition policies of the DoD, the DHS, and NASA implement some key product development principles. But they have yet to fully implement others. This gap limits agencies from ensuring a consistent approach to developing and delivering products with speed and efficiency.

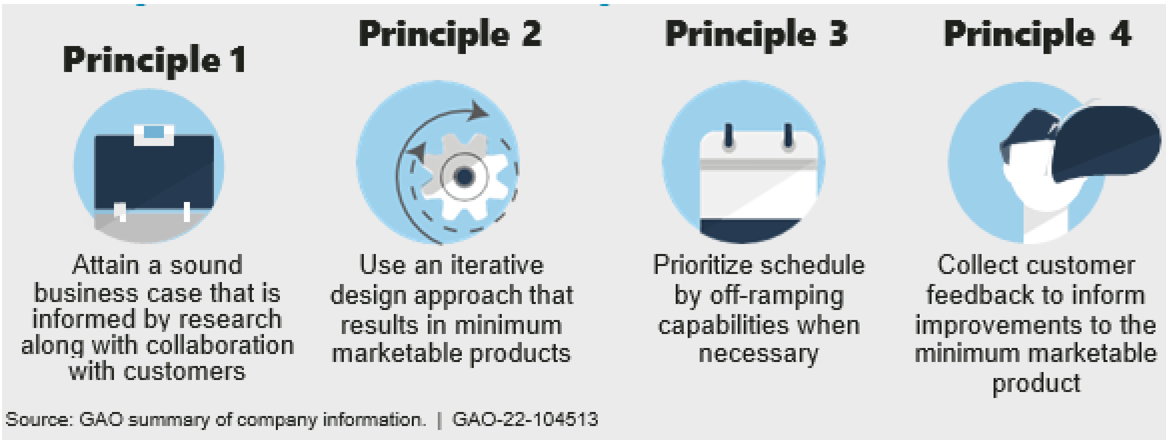


Figure 1. Leading Companies Use Four Key Principles for Product Development

For example, leading companies focus on designing a minimum marketable product—one with the minimum capabilities needed for customers to recognize value. Leading companies also prioritize a project's schedule: they release the features most critical to the customer and will off-ramp non-critical product features—an industry term for removing them from the current release—as necessary, in order to maintain schedule. Leading companies have mechanisms to solicit and implement feedback from customers early and often throughout development to ensure the product is relevant to customer needs, among other things.

What the GAO Recommends

The GAO is making nine recommendations to the DoD, the DHS, and NASA to update acquisition policies to fully implement key principles of product development. All three agencies concurred with our recommendations.

Primary DoD, DHS, and NASA acquisition policies incorporate many aspects of the four key principles, to varying degrees. However, agencies miss opportunities for positive outcomes by not addressing some sub-principles in their policies.

- The DoD's policies do not require all programs to consider off-ramping non-critical capabilities in order to achieve schedule, hindering programs' best chance of maintaining time frames.
- The DHS's policies do not require all programs to utilize modern design tools during hardware and software development, limiting consistent opportunities for programs to successfully improve revisions to the design.
- NASA's policies do not include mechanisms for programs to obtain and utilize product feedback from stakeholders or end users—such as astronauts using spacecraft or the science community benefiting from NASA projects—in order to identify challenges or new features to include in subsequent projects.

The GAO previously found that other factors beyond policies can affect agency outcomes, including structural differences between government and private industry. However, the GAO's prior work also demonstrates that key principles from private industry can be thoughtfully applied to government acquisition to improve outcomes, even with the different cultures and incentives.

How the GAO Did This Study

This report examines principles that guide leading companies' product development efforts and the extent to which primary, department-wide DoD, DHS, and NASA acquisition policies reflect the companies' key principles and result in similar outcomes. The GAO identified the 13 leading product development companies based on rankings in well-recognized lists; interviewed company representatives; analyzed department-wide acquisition policies from the DoD, the DHS, and NASA; and interviewed agency officials. The report is the first product in a planned body of work. In future work, the GAO will explore how government agencies can apply some of the key principles outlined in this report.

Open for Business: Business Models for Innovation with Modular Open Systems Approaches

Greg Sanders— is a Fellow in the International Security Program and Deputy Director of the Defense-Industrial Initiatives Group at CSIS, where he manages a research team that analyzes data on U.S. government contract spending and other budget and acquisition issues. In support of these goals, he employs SQL Server, as well as the statistical programming language R. Sanders holds a master's degree in international studies from the University of Denver, and he holds a bachelor's degree in government and politics and a bachelor's degree in computer science from the University of Maryland. [gsanders@csis.org]

Contributing Authors: Leah Matchett, Kasey Welch, Laura Bocek, Alexander Holderness

Abstract

Modular Open Systems Approaches (MOSA) build on techniques used in the commercial world to attempt to bring innovation, speed, and savings to Department of Defense (DoD) acquisition. However, while competition can be a powerful motivator, MOSA can be disruptive to those traditional defense industrial base business models that rely on the expectation of long-term production and sustainment revenue to make back corporate investments. This project undertook interviews and surveys to better understand how MOSA influences vendor incentives and what business models may best serve DoD needs going forward. MOSA's promise of enabling faster technology refresh and bringing in new sources of innovation addresses technical and operational challenges associated with 21st century great power competition and longstanding DoD difficulties in accessing commercial technology.

This project has identified three overarching challenges regarding MOSA adoption: communicating and demonstrating government commitment; developing a MOSA-enabled IP and data rights strategy; and establishing standards and interfaces. In addressing these three challenges, the government will need to employ its acquisition toolkit to take different approaches with different vendors. To better understand how to make this transition a success, this paper presents a framework for evaluating the DoD's readiness for MOSA.

Introduction

Across multiple decades, the Department of Defense has tried to emulate the advantages of commercial sector approaches (like common standards for personal computers and telecoms) in its acquisition system, using open standards to make it possible for a range of suppliers to innovate and compete with one another while still developing compatible technology. The present drive towards Modular Open Systems Approaches (MOSA) is a central pillar in this effort, one that has been repeatedly encouraged by Congress—including in the 2017 National Defense Authorization Act, which codified a requirement to use open interfaces in major defense acquisition projects. The possibility of a shift to MOSA for defense acquisition has drawn significant interest within the defense community (Baldwin, 2019; Minor, 2017). It has the potential to increase competition among vendors, make integration of subsystems and components cheaper and easier, increase interoperability, decrease the cost of operations and maintenance, and encourage innovation.

However, the potential benefits of MOSA also carry significant implications for defense–industrial business models. By making it easier to change subsystems (and the components within them), MOSA adds uncertainty to the level and duration of the business a vendor earns when selected to provide technology on a system. Successful MOSA implementation implies a greater likelihood that a vendor's technology could be replaced, or at least recompeted for, during a system's production. It also increases the ease with which a vendor's technology can be replaced in the sustainment phase. Historically, major defense contractors receive a significant portion of their revenue in the sustainment phase of a platform's life cycle. One reason for the substantial revenue from the sustainment period is that the long platform lifespans are vendor-locked. The government is often reliant on a single vendor for upgrading the system, integrating a component from a third party, or any variety of maintenance tasks. This vendor-lock can be reinforced by proprietary interfaces that make their owners a mandatory participant in upgrades and gives them control over the supply chain throughout the life cycle of a system. The transition to open standards alters vendors' ability to forecast future revenues, especially during sustainment, and thus has the potential to reduce returns on investment, undermining companies' business models. On the other hand, a MOSA environment brings new opportunities to the industrial base. Even incumbent suppliers benefit from some of the changes in a MOSA environment, as it can lower transaction costs and creates opportunities to acquire market share previously controlled by competitors (Scheurer & Moshinsky, 2020).

Prior scholarship on MOSA seeks to lay out the technical and business challenges from a theoretical perspective. However, as MOSA moves closer to being in widespread use, it is vital to understand the perspective of industry as they navigate this paradigm shift in their business model. This report seeks to bridge the divide between the theoretical framework for MOSA and the experiences of practitioners through of survey and interviews. The insights gleaned from this work provides an important contribution to the MOSA literature and empowers policy-makers with new information as they seek to better understand the MOSA problem set.

The government has a vital role in implementing MOSA but cannot succeed alone. Successful implementation of MOSA will require robust participation by a cadre of vendors who are ready to put down the initial investment to make their products MOSA-compatible and ultimately willing to invest in technology that leverages MOSA to provide the benefits the government desires. For this reason, industry's understanding of and concerns about MOSA—particularly its impact on the business model—is key to MOSA implementation. To investigate how companies think about their business models in a shift to MOSA, the authors of this report conducted surveys and a series of interviews, primarily with those in industry. While a wide range of projects was discussed, the interviewees and the surveyed population paid special attention to the pivotal Army aviation sector and the development of next-generation helicopter and tiltrotor platforms via the Future Vertical Lift program.

The current report identifies three key challenges as primary areas of focus for adopting MOSA in light of industry concerns. The first challenge is communicating commitment—including clarifying what objectives supported by MOSA are top government priorities, being prepared to stand by those priorities, and having a means of evaluating whether those priorities are being met. The second is clarifying government requirements for MOSA-enabled IP and data rights. This relates to understanding vendor concerns about the scope of government demands, while determining what core technical data is necessary and what boundaries can be set to give industry the room to profit on its investment. The third challenge is choosing interfaces and standards, with a particular concern for the commonality of interfaces across different platforms and the modularity of those interfaces. The project also explores the acquisition toolkit, which is not a direct focus of vendor concern but instead provides opportunities to address the challenges listed above.

This paper begins by reviewing what has already been established in the literature, with special attention to the benefits suppliers may see from MOSA. The paper then takes a look at previous MOSA research on the four cross-cutting categories mentioned above: communicating commitment, IP and data rights, standards and interfaces, and the acquisition toolkit.

The next sections of the paper focus on the interviews and surveys themselves. First, the methodology of the survey is described: a mix of 1-hour interviews and electronic surveys reaching a more widespread audience. Both approaches were performed with the participants' inputs handled on a "not for attribution" basis, to encourage robust commentary and participation. The organizational and individual demographics of the respondents are summarized to give the results context. The paper then walks through the results of the survey, broken down into six large categories:

- Big picture opinions: how vendors are inclined toward MOSA, how it affects their financial incentives, and how their business model could change to incorporate open systems. This section discusses a division identified between MOSA embracers and MOSA-hesitant respondents, sometimes within the same organization.
- The open interfaces marking the boundaries between different modular systems, along with their associated challenge: shaping which subsystems and components will be available for competition.
- Industry views on intellectual property and data rights: how they lie at the core of their business model and what rights the government may require in implementing MOSA. This is especially relevant for companies that sell to commercial customers.
- The government's MOSA readiness and supporting infrastructure, including workforce and investments, and to what extent these are seen as opportunities by industry.
- Specific acquisition approaches, such as other transaction authority (OTA) arrangements and licensing, and how they shape company incentives. This section also covers one specific acquisition controversy: the role of the system integrator.
- Industry perceptions of the outcomes achieved by MOSA projects and their sources of innovation. These outcomes are a key indicator of success from the government's perspective, but they can also reflect favorably on future opportunities for companies.

The paper concludes by discussing findings related to the three challenges mentioned above: communicating commitment, MOSA-enabled IP and data rights, and choosing interfaces and standards. It also investigates the potential of the acquisition toolkit to allow for diverse approaches to addressing vendor incentives.

Background and Literature Review

MOSA and related open-system architecture topics have a multidecade history within government, which involves a mix of interwoven technical and business considerations. Much of the literature focuses on the perspective of the acquirer, both commercial and government, as well as a range of policy- and technology-focused issues. This section starts by reviewing the core concepts of successful MOSA implementation, then dives deeper into five key concepts. The first topic addresses the overarching question of what motivates vendors in a MOSA environment. The second topic covers communicating commitment to an appropriate set of MOSA objectives and being able to confirm that openness has been achieved. The third topic presents the foundational choice of interfaces and standards, along with the implications that flow from those choices. The fourth topic is intellectual property and data rights, their interaction with MOSA, and the need for openness in key areas. The last topic is the acquisition toolkit that seeks to align vendor and government incentives.

Starting with the big picture, the Government Accountability Office (GAO) studied private sector open-system successes in addition to speaking with military open-system practitioners. They asked what would be needed to achieve success with open systems and put forward a few central practices and enablers. The first was the importance of "broad industry support and coordination" in the development and adoption of standards, in order to create demand for open systems and to drive competition for the development of software and hardware. The second factor was a "long-term commitment" by the acquirer to "develop, implement, test, and refine standards" (GAO, 2014, p. 2) Another tenet was ensuring that an acquirer has the "technical expertise" to identify which standards to employ and which interfaces to open. Responders also indicated that "knowledge sharing across all the departments" within an acquirer was important to win organizational resources and minimize necessary investments (GAO, 2014, p. 2).

What Motivates Vendors

The exact benefits offered by MOSA vary somewhat from source to source, but the short list includes greater competition, interoperability, easier upgrades, incorporated innovation, and savings through reuse (*DAU Acquipedia*, n.d.; Zimmerman et al., 2019). A GAO report from 2013 contrasts the benefits of MOSA to a critical depiction of typical acquisition: "Traditionally, DoD has acquired proprietary systems that limit opportunities for competition and cannot readily be upgraded because the government is locked into the original suppliers" (GAP, 2013, p. 1). When determining which incentive structure to pursue, it is important to understand why industry can be motivated to adopt MOSA despite the differences in incentive structure, particularly considering that undercutting vendor-lock could threaten certain incumbents.

Nickolas Guertin and Douglas Schmidt (2018), of Carnegie Mellon University's Software Enterprise Institute, offer three main reasons vendors are pursuing MOSA: "(1) to avoid being left behind as others find new opportunities and (2) to take advantage of new methods to improve internal corporate efficiency, as well as to (3) increase market share and increase profits."

Bob Scheurer and Ed Moshinsky (2020), co-chairs of the National Defense Industrial Association System Engineering Architecture Committee, elaborate on the efficiencies and potential sources of competitive advantage and outline seven different positive-sum benefits of adopting MOSA for suppliers:

1. More competitive products through lower cost structures
2. Faster time to market, with less development time and costs
3. Increased competition within supply chain for lower costs
4. Increased interoperability providing greater market opportunities
5. Structured upgrade paths for quicker tech refresh and longer product life spans
6. Foundation for greater commonality across products, and larger lot buys for reduced costs through modularity
7. Incentive to innovate via an improved IP policy, by allowing access to and integration of critical supplier IP while still protecting supplier business interests and investments (2020, p. 6)

Well-architected MOSA makes designing products in a complex interrelated technology ecosystem easier and enables suppliers to focus more of their efforts and resources on product quality. In addition, as point 3 above implies, suppliers often act as acquirers themselves and thus have the potential to gain some of the benefits that the DoD seeks. The points on interoperability and commonality indicate that scale is another potential supplier benefit, as the promulgation of open standards means that a product may be useful to a greater range of customers while reducing the need for modification. The potential for greater product lifespans also reduces the negative impacts for suppliers that face greater competition. If incumbent companies can more easily refresh the technology in their product, then they can potentially steadily improve existing product lines rather than sell the same system for longer.[53]

Taken together, the two lists above show why transition to MOSA can easily build momentum—or falter due to insufficient adaptation. The efficiencies cited by Scheurer and Moshinsky (2020) become more widespread when an acquirer makes greater use of MOSA; this is also the case if multiple buyers (e.g., different military departments or allied countries) choose the same interoperable standards. Meanwhile, Guertin and Schmidt's (2018) first point suggests that this is a transition for a larger sector of vendors, and that even a vendor that finds MOSA less appealing may still adopt lest they lose access to the new opportunities that their competitors are able to exploit. Thus, the benefits of MOSA for suppliers can create a virtuous cycle, but the uncertainty inherent when implementation is first starting out can delay or potentially short-circuit the development of that cycle.

Some suppliers will seek advantage by being early adopters of MOSA. Because the DoD often employs outside vendors directly in the creation of architectures and standards and in the management of integration, these early adopters have an opportunity to distinguish themselves from competitors by offering more openness in the initial design and implementation of MOSA. Davendralingam et al. (2019) highlight the Army VICTORY program, noting "the participation of GE Intelligent Platforms (Charlottesville, VA), which supported the use of an open standards approach, seeing it as a key business opportunity since other prime vendors were focused on proprietary-based solutions"(p. 393).

Survey Methodology and Respondent Demographics

The industry opinions in this project are based on a series of not-for-attribution interviews supplemented by an electronic survey sent to interviewees and to the membership of the Vertical Lift Consortium. The interviews were targeted at companies that have experience with MOSA and were reviewed by the Army to ensure that a selection of companies they regularly work with would have ample opportunity to participate. A total of 16 guided interviews lasting roughly 1 hour were conducted, some with multiple participants. Interviewees sometimes included different divisions or experts within the same larger company; in total, 10 vendors participated. The interview process also included speaking with analysts and government practitioners, and a small number of international experts were included on both the vendor and government sides.

The project further developed an extensive survey to elicit respondents' opinions on how MOSA will affect their individual work and their company's business model. This survey took two forms: a long form for those who had a "nuanced understanding" of MOSA, and a short form for those who were only generally aware of the concept. The short survey focused on the following areas: respondent and company characteristics, impressions of MOSA overall, impression of steps needed for successful MOSA implementation, and the relative importance of various MOSA initiatives to the companies. The long form for those with specific MOSA familiarity asked all of these questions but also inquired about their thoughts on how intellectual property and

[53] The NDIA suggests that a carefully balanced version of MOSA will be most successful, and that, as per the information in their report, "the NDIA expects that all stakeholders in a MOSA implementation can achieve a higher potential for success and realize both the technical and business benefits from such implementations on system development programs and deployments" (2020, p. 6).

data rights and interfaces have worked in the past, as well as how MOSA affects their company's acquisition approach, possibility for innovation, financial incentives, and outcomes. For the ordered multiple-choice questions that are displayed on a Likert scale below, no default answer was specified, but the most negative answer in a given context was listed first. Two multiple-choice and most short-answer questions were optional, enabling some respondents to skip these questions.

The survey was sent to all interview respondents, as well as to the Vertical Lift Consortium mailing list. In total, 13 responses to the short form survey and 50 responses to the long form survey were collected. The short form survey was a subset of the longform survey so those questions on the short form survey were answered by both all that completed the survey. Of the 63 completed electronic survey responses, roughly a quarter were completed by interview respondents, with the remainder coming from the Vertical Lift Consortium.[54]

Respondent Demographics

Given the diversity within the defense industrial base, the survey started by asking the respondents to self-classify their organization and their work.[55] Multiple respondents from the same organization—including organizations that separately participated in the interview—were allowed, even if at times classification of the organization varied between different people at the same organization. As shown in Figure 1, the two roles most represented are original equipment manufacturers (OEMs) and system integrators, who make up over half of the respondents. These categories of organizations (sometimes in partnership with one another) are the traditional platform leading primes.

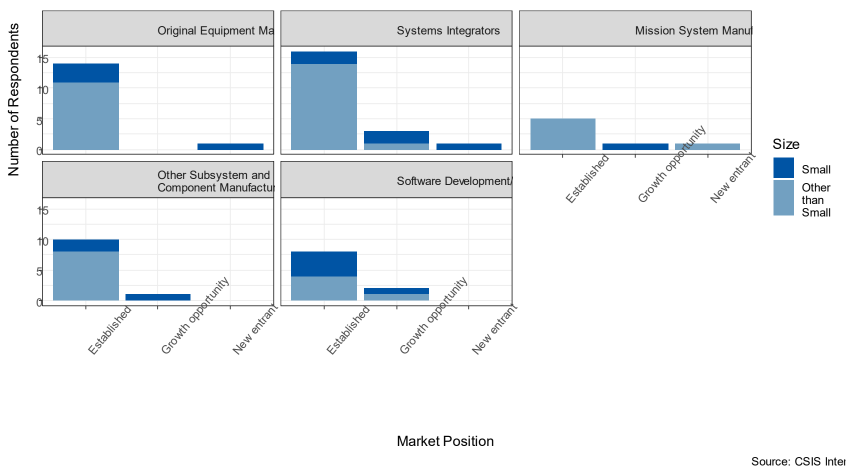


Figure 8. Industry Survey Participants Self-Classification by Organization Role and Market Position

The other three categories are key to providing modules to the platform, although vendors for complex mission systems and subsystems often have their own integration responsibilities within their domain. Most of the respondents were established companies, although there were three new entrants and seven respondents seeing substantial growth

[54] Approximately 40 additional respondents started but did not complete the survey.
[55] Analysts were asked to describe the organizations with which they were most familiar.

opportunities. The relative frequency of established players partially reflects an orientation of interviews to MOSA-experienced companies, but also suggests that even smaller vendors in consortiums often see themselves as established players. A significant minority of respondents (28.6%) were from small businesses, partly balancing the low proportion of non-traditional defense companies. Collectively, this means that the results will better capture the range of opinions among larger prime competitors than smaller or less traditional competitors. Thus, discussion of adopting commercial technology will often refer to the commercial portion of an established defense industrial base company or to the challenges of bringing in a third party's parts and services.

The majority of respondents included aircraft in their focus, while a near-majority also included electronics, comms, and sensors. This again reflects the focus on FVL-relevant vendors and the Vertical Lift Consortium, but it does show that respondents have a wide range of focus areas. While the interviewees (and the respondents writ large) most frequently were focused on FVL and Army programs more broadly, cited experience with MOSA projects included all three military departments.

Looking at the individual responsibility and expertise of respondents, engineers and program managers were the most common respondents, with government relations and contracting officials being a distant third. The two analysts were experts that study MOSA programs and industry rather than being vendors themselves. In their individual responsibilities, over 60% of the respondents dealt with an even mix of hardware and software, and among the remaining, more dealt with hardware than software. This sample has extensive knowledge of MOSA, with nearly four fifths having at least a nuanced understanding and more than half of respondents having direct professional experience or outright expertise.

Frameworks and Crosscutting Patterns

This report suggests several key challenges that MOSA presents for the business case of a mix of suppliers in the defense industry. This section outlines how these challenges are related to the Army's stated goals for MOSA, how these goals are related, and how a determination might be made across acquisition priorities. It concludes with recommendations to address each challenge in the report.

The areas of focus identified in the report are communicating commitment to MOSA, developing an IP strategy that benefits from MOSA, and achieving commonality of standards and interfaces. Each of these challenges has been discussed at length in the above sections, but is briefly summarized again below:

- **Communicating Commitment to MOSA.** Companies communicated multiple related concerns:
    - Communication: Vendors, especially the MOSA hesitant, want to know the government's priorities regarding MOSA, in order to guide investments and to build the case for tailored IP and data rights.
    - Commitment: For MOSA embracers, a chief concern is that the government will accept defections from stated MOSA commitments, undercutting their process reforms. For both embracers and the hesitant, consistency between leadership's vision for the "big idea" of MOSA and the staff officers managing implementation is a key prerequisite for achieving lasting change.
    - Follow-Through: The government will have to effectively judge compliance with standards at the start and throughout the life of a program. This can involve

competition to judge integration speed, cross-vendor feedback mechanisms, and accessible conformance testing facilities.

- **IP and Data Rights.** Companies expressed concern that the government "wants it all" regarding IP and data, including commercial IP and IP stemming from vendor investments. In many cases, flexibility on IP below the interface level would address these concerns. Business-model impacts for interfaces and architectures are more challenging. Deeper interfaces implicate more sensitive data rights.
- **Standards and Interfaces:** Companies had different preferences over the extent of modularity, with some seeing disruption to existing products and business approaches and others seeing new opportunities for competition and innovation. Regardless of the extent of modularity, cross-platform commonality is a key incentive with substantial room for improvement.

A fourth cross-cutting topic is the **acquisition toolkit**. This is not a challenge in the same manner as the other three, in that it was not a direct source of concern; instead, applying a diverse mix of contracting approaches has the potential to help tune business models to address these challenges.

Table 1 connects each of the key challenges laid out in this report to larger problems the Army seeks to address. The three middle columns in the table each represent one of the challenges to successful MOSA implementation. The rightmost column covers options provided by the acquisition toolkit. The rows of the table represent ways to address the cross-cutting topic that would also aid in addressing larger Army goals. Because not every challenge will affect every goal, and because these goals may trade off against each other, this table allows for prioritization across the MOSA challenges laid out in this report, based on which goal the Army wants to prioritize. For example, if the Army's priority is accelerating development timelines, this analysis suggests that IP and data rights will not have a large impact on this goal, but that communicating commitment by incentivizing faster integration and achieving commonality in standards and interfaces will make a bigger difference.

Table 2. Connection between Army Goals and Key Business Case Challenges Identified in This Report

Army Goals	**Communicating Commitment to MOSA:** *Successfully communicating commitment to MOSA means...*	**IP and Data Rights:** *Successfully implementing a MOSA-enabled IP and data rights strategy means...*	**Standards and Interfaces:** *Successfully implementing standards and interfaces means…*	**Acquisition Toolkit:** *The acquisition toolkit can aid in addressing these challenges by...*
Design a system architecture	Convincing involved actors that the Army will not change the system later to decrease prices in the short term. Achieving a common understanding of underlying goals across both government and industry; reinforced by evaluations and testing approaches.	Determining necessary IP and data rights for technical baseline; being prepared for upfront prices to acquire them.	Encouraging cross-platform adoption. Deciding which features of MOSA and other acquisition priorities are most important, to guide choice of architecture depth. ***Key Trade-off: Extent of Modularization*	Exploring contracting approaches that consider openness when determining what proposals offer the best value. More experimentally, exploring rewarding future integration successes and wider reuse of interface, including licenses and royalties.

Army Goals	Communicating Commitment to MOSA: *Successfully communicating commitment to MOSA means...*	IP and Data Rights: *Successfully implementing a MOSA-enabled IP and data rights strategy means...*	Standards and Interfaces: *Successfully implementing standards and interfaces means…*	Acquisition Toolkit: *The acquisition toolkit can aid in addressing these challenges by...*
Develop new capabilities	Sharing a roadmap for future system development through working groups.	Encouraging investment by allowing industry to retain below-interface IP, while preserving open interfaces.	Enabling new and growing vendors to offer and be competitive in providing new capabilities. ***Key Trade-off: Extent of Modularization*	Allowing for more iterative development of requirements. Employing rapid contracting approaches, especially for software.
Maintain stable budgets for mission system development and deployment	Taking enabling measures to ensure that openness is maintained over time.	Lowering risk for government and vendors by allowing more tailored IP solutions while preserving options for future competition.	Setting realistic expectations for industry to support investing in capability development.	Enabling acquisition tool planning that enables successful budget execution. Balancing upfront costs and life-cycle costs through approaches such as minimum orders or commercial licensing.
Accelerate development timelines	Employing competitive measures that test integration speed and incentive openness on the integrator side.	Maturing tailored data rights requirements for more rapid agreements with industry and clarity in expectations.	Achieving greater commonality in interfaces or adherence to commercial standards.	Lowering time to contract, especially for software development.
Address fluctuations and uncertainties in order quantities	Setting priorities across the entire system to ensure that initial requirements are credible.	Addressing “can I still sell this” industry concerns beyond the initial platform.	Adopting standards widely, and across platforms, to mitigate investment risk through larger sales opportunities.	Diversifying acquisition approaches using ones that address risk tolerance, such as minimum order and time options.
Incorporate Software	Establishing a conformance process that is clear and viable, making it easier and less costly to integrate software.	Developing licenses that give access for key government purposes, including ease of replacement and cybersecurity, while addressing industry concerns that source code may be transferred to competitors.	Creating a common digital backbone that should improve the ability to incorporate software. ***Key Trade-off: Extent of Modularization*	Licensing software and easing the shift for vendors traditionally reliant on hardware sales.

Table 3 MOSA Readiness Framework

Category	Dimension	High Readiness	Low Readiness
Communication and Commitment	Government and broad industry communication about achievable first iterations and future roadmaps	• MOSA part of system engineering • Informed by industry input	• Lack of clarity on which MOSA goals are integral to system • Government and industry talk past each other
	Sustained government commitment to MOSA objectives and a credible MOSA funding model	• Competition by best value and contract incentives tied to openness • Budgets support iteration within a program and cross-program investments	• Government awarding contracts that undercut MOSA standards for other benefits • Absence of metrics for MOSA goals in execution
Enabling Environment	DOD and service enterprise investments in MOSA enablers	• Support from above • Encouraging adoption of widespread open standards	• Programs face MOSA challenges alone • Initiating new bespoke approaches
	Government engagement with key enablers and cross-platform standards development that builds and sustains consensus with industry	• Wide range of industry stakeholders engaging in the standards development process	• Limited interaction between industry and government • A few vendors dominate the process at the expense of others
	Acquisition and sustainment workforces' business and technical expertise	• Widespread understanding backed by effectively deployed experts track larger MOSA concepts and meet MOSA goals	• Experts are rare within organizations • Solutions are applied that run contrary to leadership direction
Business Models	Business models that incentivize defense-industrial base transition	• Diverse mix of contracting approaches that use a variety of incentives to meet MOSA goals	• Contracting approaches that disproportionately rely on significant profits during the sustainment period
	Expansion of the supplier base and inclusion of commercial technology	• New vendors competing for modules • Reduction in the bifurcation between the military and commercial markets	• Low level of engagement in DoD standards process • Lack of knowledge or interest in contracting opportunities

The solutions to these three MOSA challenges and the employment of the acquisition toolkit are not independent. Picking an interface standard is unlikely to have any influence on outcomes if the commitment and follow-through measures are not sufficient to ensure that the implementation is open in practice. Likewise, failing to tailor IP approaches will almost preclude successfully finding ways to incorporate commercial technology. The depth of interfaces should be shaped by larger MOSA and system objectives, such as the use of multifunction shared computing resources, and it in turn will shape what IP and data rights are necessary for the architecture. Addressing any one of the three challenges well will make the others easier, but entirely neglecting one will undercut attempts to address the others. The final section of the report outlines several ways that the government might do this. However, even if these requirements are met, many of the Army's stated goals for MOSA will falter if the challenges of commitment, conformance, and incorporating technology are not addressed as well.

MOSA Readiness Framework

In a parallel effort to the industrial survey covered in this report, the authors created MOSA readiness framework shown in Table 2 (Sanders & Holderness, 2021). This framework suggests metrics for considering the government's readiness for MOSA's coordination problems.

Discussion

Communicating Commitment

Open system adoption is a coordination problem offering shared benefits for government and industry, but it also carries transition costs and risks. "Communication" here is used broadly—strategic outreach and conferences are directly relevant, but the bigger picture is demonstrably taking steps that demonstrate both commitment and the capability to follow through. Effective coordination has multiple aspects: vendor participation in standards, adoption of interfaces, and implementation of open-source business process reforms; vendors investing in and proposing technologies of interest; vendors providing feedback and sharing knowledge to shape effective approaches; and vendors being discouraged from behavior and proposals that would fail to achieve openness.

Vendors, especially MOSA-hesitant ones, desire to know the government's MOSA-related priorities for them to guide their investments and build their case for tailored IP and data rights arrangements. There is some risk here that defining goals and priorities too narrowly may foreclose future options that are compatible with larger MOSA principles but are not the focus of today's leaders. However, that risk is balanced by the opportunity to show how MOSA goals fit into the larger program and enterprise goals, and how they will stick around even when trade-offs must be made in development.

Use cases are one such mechanism for communication. These are scenarios that depict how the MOSA characteristics of a project are to come into play to achieve desired goals. Moving from a broad goal, like providing competition and technology insertion, to a use-case example of how a new mission system module may be incorporated provides more detail for vendors, but it remains goal-focused in a way that does not need to be highly prescriptive to be effectively communicated.

For MOSA embracers, a chief concern is that the government will fail to enforce MOSA. During the present FVL competition phase, this could mean accepting a proposal that offered a lower front-end price but fell short on openness goals. Further on in the life of a system, this might mean accepting an exciting module that withholds key information or otherwise fails to conform. In either case, accepting a solution that is putatively "close enough" could undermine the openness of a system in ways that manifest over time in exchange for a short-term benefit. This fear has some basis in the multidecade history of open system policy goals, which have often lost out to proprietary systems in practice.

Those that raised this concern emphasized its importance to their incentives, but in broader terms they had positive views of Army commitment. Upfront work on standards, and steps such as the creation of the Architecture Collaborative Working Group, were seen as important sign of intended follow-through. Excess detail in requirements was widely agreed to be a risk, but suggestions as to what was necessary varied, with some arguing for mandating chosen standards and others suggesting that specifying the open interfaces was more important.

Beyond the MOSA embracers, a wider pool of vendors underlined the importance of commitment to the programs that contain MOSA. The logic here is straightforward: when a program is delayed and shrunk—or worse, canceled—many of the investments made in it will see greatly reduced returns. The success of the larger efforts of transitioning to MOSA depends in no small part on the success of the present crop of programs that are mandated to implement it.

There are opportunities for industry to demonstrate its ability to implement MOSA goals beyond thresholds. Setting threshold requirements to encourage openness is part of the incentive picture, but many aspects of openness will not be pass-fail. Developing metrics to measure MOSA alignment is highlighted in the literature as a prerequisite for setting appropriate incentives for vendors. In terms of competitive evaluation, some MOSA embracers suggested that past performance and best-value criteria for openness could be used to set apart those that could deliver an architecture that would provide a greater savings in the long term.

Any metrics and incentives chosen will come with some controversy and vendor feedback. For example, past performance may be less applicable to non-traditional vendors or fail to account for decisions on the government side. Nonetheless, one promising approach emerged during the interviews. New hardware and software integration will regularly be a key MOSA priority and one which a broad range of primes express confidence in their ability to execute. When prototypes are sufficiently advanced and developed standards are in place, integration "shoot-offs" are one way to put this confidence to the test. Under this approach, the government would furnish software or hardware products that conform to the pertinent open standards and give the relevant integrator the opportunity to demonstrate how quickly and effectively these can be incorporated into their system. One system integrator did warn that some level of communication between the integrator and the module provider may be necessary.

Sustaining commitment does not end when winning vendors are chosen, but instead it should be tracked throughout the life of a program. Testing the speed of integration has value post-competition as a means of evaluating the sufficiency of openness and available technical data and artifacts. In one international example, an architecture intended for wide deployment is being tested by a third party taking on the role of integrator. In this particular example, the vendor committed to make additional artifacts available if the previous IP and data rights scope was insufficient to enable integration. Specially-negotiated license arrangements may benefit from similarly being tied to goals rather than to static predictions of what IP and data rights would be required.

IP and Data Rights

For a plurality of vendors, IP and data rights are the top concern regarding MOSA. The incentives are most pertinent when vendors are bringing technology they also sell in the commercial market, with regards to front-end investments, and for sustaining investment. The open interfaces themselves are only part of the discussion, which expands to grapple with the question of the contours of vendor IP under MOSA and what incentives this creates for investments.

A common industry concern is that government "wants it all," even IP developed at private expense or when not in support of a clear goal. Vendors often argued that accommodation could be reached on a range of more limited transfers, with in-field maintenance and depot arrangements as commonly cited examples. This complaint precedes

MOSA, but the switch to open interfaces and the rearranging of existing business models brings it to the fore. Concerns were diffuse, but two areas received repeated attention: transfer of IP to competitors and computer source code. Transfer to a competitor might happen intentionally (for example, as part of substituting a new system integrator for sustainment) or unintentionally (as a consequence of greater openness) if not accompanied by protective measures. The government's desire for computer source code, without being accompanied by a license, was repeatedly raised in the electronic survey.

Especially for commercial products, industry wants to know "can I still sell this?" MOSA brings opportunities to adapt existing technologies to a new system, although this can raise questions about whether the ability to sell the underlying technology is affected, especially when the government is paying for the adaptation. For example, would a commercial avionics package adapted for use in a military helicopter come under export control regulations? For software sales, one vendor raised the question of whether a sale would be a one-off and afterwards available to any part of the government.

MOSA has the potential to enable a diverse mix of IP and data rights approaches. Building on an approach proposed by Guertin and Schmidt's (2018) framework, it is possible to determine what IP and data rights are necessary by thinking of the system in two tiers. The first tier is the fundamental necessities in the architecture, standards, and infrastructure that are the foundation of MOSA. Industry does have concerns—sometimes vehement ones—in this area that will be challenging to finesse. Clear explanations of goals and developing plans that will lead to cross-platform adoption has aided this step in one international case. However, wise choices and a willingness to bear some upfront costs will be critical.

With that foundation in place, MOSA can then enable a diverse mix of IP and data rights strategies for the replaceable portions of system. These modules will be "grey boxes," not truly opaque black boxes, as information about their workings will be required for successful integration. However, the option to turn to a new vendor means that as long as any module conforms with the larger standards, there is much more room to reach a range of IP and data rights approaches that are appropriate to the vendor, the need they are meeting, and the mix of investment. A range of vendors independently raised this point in tandem with their IP and data rights concerns. This is the realm where acquisition approaches, such as software licenses, that give vendors something to hold on to and that reward investments can be experimented with at lower risk of future vendor-lock. The MOSA goals the government seeks for a project may change and evolve over time in ways that cannot be predicted in advance, even by the best tailored arrangement, but the availability of competition means that departing from arrangements that are no longer suitable is a viable option.

Business model impacts for interfaces and architectures are more challenging. Deeper interfaces implicate more sensitive data rights. While MOSA is quite compatible with leaving "grey boxes" preserved for vendor investment, the size of those boxes depends on the choices of interface and standards. What might be a single package under one architecture could be broken into multiple components in another, or divided between hardware and software in yet another. Under deeper interfaces, what once was internal to a subsystem may now openly flow across components, including those made by different vendors. For example, under CMOSS, a box containing a radio and antenna would be broken up, with key capabilities placed on hardware cards and the antenna itself used as a pooled resource serving multiple functions (Strout, 2021). However, for a mission system manufacturer, this might mean that offering a commercial product would be unappealing, as it would involve breaking up a subsystem sold to the larger market as a whole. Likewise, commercial technology—notably software—may have special licensing requirements that can be preserved when they are part of a large subsystem but which may not be compatible with them being a module in their own right.

On the other hand, more granular interfaces can also open up opportunities for a range of vendors, including software developers and component manufacturers. While special licensing or other acquisition approaches may be necessary for some non-traditional vendors, these should be compatible with the MOSA-enabled IP and data rights diversity discussed above.

Standards and Interfaces

Depth of interfaces is a dividing line. More granular interfaces may bring benefits for multifunctionality and more advanced components, but the MOSA-hesitant see integration risks and a loss of incentive to invest. Carrying over from the last point under IP and data rights, the extent of the advantages of greater modularization was a point of disagreement. For some vendors and other interviewed experts, more modular interface approaches are desirable, as they allow for greater multifunctionality and more rapid insertion of technologies such as sensors or processors rather than tying them to the refresh cycles of the larger subsystem. Multifunction components have the potential to reduce size, weight, and power (SWaP) by reducing duplication. For some vendors, including ones with mixed feelings on MOSA, a more modular architecture may be their best chance to be competitive for a system where an established vendor has been taking a leading role on the subsystem in question. For the skeptics, including many MOSA-hesitant respondents, more granular interfaces did raise concerns about IP and data rights and their incentives to invest, but other concerns were also cited. Namely, a more modular system does mean a greater integration role for some mix of the system integrator and the government. One related concern is the need to be clear about who is responsible for the performance of different parts of the system when a failure comes out of interactions rather than a single module.

Widely adopted standards and interfaces expand the potential market and are a key incentive. When a technology works with one platform employing a common standard, MOSA greatly speeds the development time and lowers the cost of bringing that technology to a new platform with compatible standards. As vendors incorporate a standard into their businesses processes, future opportunities to employ it are also made easier. This commonality also can result in a module having a diverse portfolio of potential customers and thus being less affected by volatility in any individual program. Multiple respondents, including those expressing hesitancy about the downsides of more granular standards, noted that market size was a key positive incentive. This was not a universal sentiment; some standard choices or implementation will bring controversy, but there is a clear upside to increasing standard reuse, which was reinforced in the vendor survey regarding the anticipated advantages to vendors of allowing part reuse.

Standard-setting bodies and other coordinating groups are vital feedback mechanisms and benefit from openness with allied countries. Standards commonality could put the brakes on the ability to incorporate new technology if the standards were static. Happily, the bodies instituted for industry to discuss standards, as well as other coordinating mechanisms such as software interface control working groups, were seen as venues both for shaping future developments and for the government to provide roadmaps for future intentions. The diversity of approaches in industry provides an opportunity to vet ideas that may run the risk of undermining openness (even if they would be advantageous to a company) by bringing them to bodies like the Architecture Collaborative Working Group. One challenge raised by some international interviewees was that some standards are partially classified, and that some consortiums can limit participation only to the U.S. subsidiaries of international companies. The FACE standard does better in this area than some other DoD standards, but this is an area where classification and no-foreign limitations should be used sparingly, as allied adoption of standards can be advantageous for interoperability and for exports—besides which, cross-national embrace of standards further increases their potential for reuse.

Adopting or following commercial open standards where possible is desirable but requires upfront thinking on cybersecurity. Some commercial standards, such as ethernet and the defense industry–focused OpenVPX, were given as examples already in use. The automotive industry was favorably cited for its success in adopting open systems; in the aerospace sector, some respondents, especially MOSA embracers and analysts, cited open integrated modular avionics architecture and efforts in Europe as models worth greater study and adoption. Adoption where possible and hewing closely when wholesale adoption was not possible was seen as a worthy goal, but also one that would require deliberate effort to implement and that would face complications—such as the risk of obsolescence, as industry sometimes switches standards entirely. Additionally, one skeptic on this idea raised concerns that standard update cycles might mean that attempts at partial adoption would just lag perpetually behind. Cybersecurity, also raised outside of the commercial standards context, was cited as an important front-end issue rather than something that could be added later.

Enterprise mechanisms to build commonality are a good investment. The most important benefits of MOSA can vary from one program to another. However, the interest of individual programs can be in conflict with that of the larger enterprise. Upfront decisions need not only consider prices years in the future, but also questions of commonality. As more programs settle on a single standard, the potential benefits and incentives for industry accrue. However, particularly for legacy programs, the costs and compromises of adopting a potentially widespread approach may make bespoke approaches win out.

While vendors will have strong opinions about what standards and approaches will win out, the advantages of working at the enterprise level were uncontroversial. The survey showed enterprise-wide governance and policy was the most favored of the options presented. Here, approaches by Army PEO Aviation garnered some praise. Likewise, cross-service coordination was seen as desirable but more of an aspiration than a topic where progress was being made.

Assigning responsibility for integration and considering airworthiness certification is worth upfront attention. A range of vendors, especially system integrators, argued that a set of conforming modules was not seen as sufficient to ensure a successful system. End-to-end integration, especially during development, was widely seen as a coordination-intensive challenge and one that would require trade-offs that go beyond enforcing and updating standards. Determining who is responsible for the performance of different parts of the system as well as the extent of government integration responsibility will be important for ensuring that accountability and authority are aligned. As a related issue, airworthiness certification and approaches that enable updates that do not endanger the larger certification are also worth early consideration. One analyst argued that incremental certification, as employed in open integrated modular avionics, is a plausible path forward.[56]

Acquisition Toolkit

The acquisition toolkit is different than the prior categories of findings because it is more contextual and less shaped by the transition to MOSA. Many of the considerations raised apply

---

[56] For example, see Daniel P. Schrage and William Lewis, "It Is Time for Army Aviation to Move to a Development Assurance Approach for Including Open Integrated Modular Avionics" (presented at the Vertical Flight Society's 76th Annual Forum & Technology Display, Virginia Beach, VA, 2020), 12.

to larger adaptable acquisition concerns, in particular for software acquisition, although proper use of the acquisition toolkit has the potential to help address the aforementioned challenges.

The contracting approach often has an "indirect" relationship with MOSA, and traditional industry is comfortable working with a range of mechanisms. While other transaction authority and single-award indefinite delivery vehicles were both favorably rated, most interviewees did not emphasize any particular mechanism as necessary. Instead, for those that spoke to considerations such as cost-based or fixed-price contracting, the emphasis was often placed on the phase of contracting and the certainty of the requirements at that time.

Contracting may be more central for non-traditional participants. Fixed-price approaches are more associated with examples of commercial technology adaption, suggesting that dynamics for non-traditional vendors in the other parts of the acquisition system apply here as well. Likewise, other transaction authority (OTA) approaches were supported but not universally favored. For traditional vendors, OTAs do bring the potential for greater speed and flexibility, but also often involve cost-sharing requirements.

Software licensing deserves more attention and use, especially for source code access, but will bring complexities. Licensing arrangements that pay for regularly updated software as a service, rather than traditional waterfall software acquisition, can be well suited to an iterative approach, with ongoing investments after reaching initial capacity. Respondents noted that this approach was applied in the commercial sector, often by their own companies. Particularly if the IP and data rights largely rested with the government, vendors argued that the incentives to make ongoing investments would be reduced. That said, licensing will require developing additional government expertise, and commercial licenses are often explicitly not designed for military use. In addition, the IP and data rights aspect of licensing involves complexities separate from payment structure, as the government may need sufficient access to trace faults while limiting what source code could be seen by competitors. In parallel, internal government software capacity may bring additional options—if also greater responsibility and talent demands on the government side. To some MOSA-hesitant vendors, depots are less of an IP and data rights concern and diverse acquisition models may include open-source approaches alongside greater use of commercial style.

Diversity of methods can help round out restraints in other areas. Choices regarding standards and interfaces benefit from cross-platform adoption, and while different sorts of platforms and mission sets may benefit from different priorities and metrics, this tuning should be weighed against commonality. As a result, while the extent of modularity may differ across platform areas and there will be some variety from one system to the next, standards and interfaces will be shaped by higher-level choices and will be a better fit for some vendors than for others. While the requirements set by an open interface apply to all vendors using it, greater diversity is possible for IP and data rights arrangements, dramatically so within modules. That said, work such as system integration, the design of open interfaces, and the open aspects of modules themselves offers less flexibility. Specially-negotiated licenses may allow for some tailoring in those areas, but maintaining commitment to the openness of the system is vital. The acquisition toolkit may be of lesser importance to the more challenging factors, but it also provides an opportunity for tuning when other aspects are fixed.

This diversity of approach suggested that extra acquisition workforce attention should be devoted to the hard cases, in particular system integration, where MOSA is changing both the business model and the nature of the responsibility. One possibility raised in interviews and at the Acquisition Research Symposium would be a royalty model where, for example, successful integration of additional vendors and technologies could reward the developer of an open interface even though they did not retain an IP interest (Tate, 2021).[57] In a follow-on survey, royalties and similar licensing approaches were widely thought to be both relevant to and to enhance the appeal of MOSA projects. More conventionally, the literature emphasized the development of metrics based on MOSA goals and tying contract incentives to these metrics. Within arrangements such as IDIQ vehicles, success might be tied with time extensions rather than be viewed as a strictly financial benefit. That said, some vendors interviewed were more worried about downside risk, especially in cases where order numbers may fluctuate for reasons that they have little influence over. When vendors are bringing key commercial or other heavily invested technology, it may be appropriate to offer minimum duration or quantity guarantees, especially where technology refresh cycles are slower. In follow-on interviews, such guarantees were supported, but not seen as a panacea. Vendors may be able to offer lower unit costs based on cost curves achievable with the guaranteed quantities, but may be unwilling to include further savings based on unit counts they are at risk of not selling.

Conclusions

MOSA Changes to the Business Model

Efforts to transition to MOSA are longstanding, but urgency is growing due to the mounting difficulty of high technology great power competition reinforced by congressional mandates. The start of a major program, such as the Army's Future Vertical Lift, represents an irreplaceable opportunity to apply standards and open interfaces. However, MOSA implementation faces a range of difficult business and technical choices and is only part of a larger industrial base management and program development picture. This report assists policy-makers and analysts by documenting the sometimes clashing views of industry, aiming to inform the development of business models that will be able to align incentives between DoD priorities and vendor interests. Success means that government and industry solve a coordination problem; this will be iterative and increase the adaptability of individual programs and the acquisition system as whole. Ongoing communication and feedback will be necessary to sustain progress, and this report assists that process by helping customers and vendors to understand one another—and to understand the prerequisites of transitioning to a system that, while still competitive, offers new opportunities and efficiencies for those eager to innovate.

To successfully carry out a transition to MOSA, the Army will face three parallel but interlinked challenges: communicating commitment, developing a framework for IP and data rights, and choosing and building consensus on standards and interfaces. Collectively, addressing these challenges will require understanding the government's own priorities in implementing MOSA, building the government's MOSA infrastructure in the form of MOSA-savvy personnel and MOSA resources, making prudent choices on IP, establishing standards and interfaces informed by a range of industry input in an ongoing dialogue, and demonstrating a willingness to bear short-term cost to achieve longer-term gains.

There is a range of industry opinions on MOSA that vary by industry position and function, sometimes even among different business units within diversified companies. Even within those traditional defense industry players most comfortable with the DoD's existing approach to system design, there are a mix of opinions about MOSA. It is clear that there is a

57 Comment at the 48-minute mark. Tate also proposed a system of two proposals, one with vendor-preferred data rights and one with government-purpose rights, to better understand the underlying problem and incentives needed.

critical mass of support for MOSA in industry, enough to make success achievable. However, there is also enough MOSA hesitancy to undermine the effort if the government makes poor implementation choices or loses its commitment to the effort. Overall, this mix of genuine MOSA support from some traditional industry players and the potential for inclusion of non-traditional players through MOSA does present an opportunity to use competition to encourage MOSA adoption. Sustaining and communicating commitment is critical to ensuring that threshold levels of openness are achieved and maintained. Ultimately, MOSA implementation will lead to a substantial reshaping of defense supply chains that will necessarily evolve over time. Addressing intellectual property and data rights is not just a matter of resolving all of these issues upfront, but of taking advantage of the benefits offered by MOSA to ensure that industry has appealing business models in a more competitive MOSA environment. Finally, the choice and development of standards and interfaces will require difficult trade-offs and will benefit from achieving commonality both within portfolios, such as PEO Aviation, but also throughout the DoD and beyond. Across these challenges, the acquisition toolkit can assist in making the transition to MOSA appealing—both through the application of traditional acquisition judgment and through the application of adaptable approaches that aid in aligning the incentives of a variety of businesses with those of the government.

References

Baldwin, K. (2019). Journal of Defense Modeling and Simulation (JDMS) special issue: Transforming the engineering enterprise—Applications of Digital Engineering and Modular Open Systems Approach. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, *16*(4), 323–324. https://doi.org/10.1177/1548512917751964

DAU Acquipedia: Modular Open Systems Approach (MOSA). (n.d.). Retrieved May 14, 2021, from https://www.dau.edu/acquipedia/pages/ArticleContent.aspx?itemid=143

Davendralingam, N., Guariniello, C., Tamaskar, S., DeLaurentis, D., & Kerman, M. (2019). Modularity research to guide MOSA implementation. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, *16*(4), 389–401. https://doi.org/10.1177/1548512917749358

GAO. (2013). *DOD efforts to adopt open systems for its unmanned aircraft systems have progressed slowly*. https://www.gao.gov/assets/660/656419.pdf

GAO. (2014). *Defense acquisitions: Review of private industry and Department of Defense open systems experiences*. https://www.gao.gov/products/gao-14-617r

Guertin, N., & Schmidt, D. (2018, October 15). Emerging opportunities in modularity and open systems architectures—First in a series. *SEI Blog*. https://insights.sei.cmu.edu/blog/emerging-opportunities-in-modularity-and-open-systems-architectures-first-in-a-series/

Minor, P. (2017). *Identifying the return on investment for Army migration to a Modular Open Systems Approach for future and legacy systems* (p. 47). The Defense Acquisition University. https://apps.dtic.mil/dtic/tr/fulltext/u2/1040338.pdf

Sanders, G., & Holderness, A. (2021). *Readiness for open systems: How prepared are the Pentagon and the defense industry to coordinate?* (p. 14). Center for Strategic and International Studies. https://www.csis.org/analysis/readiness-open-systems-how-prepared-are-pentagon-and-defense-industry-coordinate

Scheurer, B., & Moshinsky, E. (2020). *Modular Open Systems approach: Considerations impacting both acquirer and supplier adoption*. National Defense Industrial Association Systems Engineering Architecture Committee. https://www.ndia.org/-/media/sites/ndia/divisions/systems-engineering/division-papers/ndia-mosa-white-paper-final-release--ndia-architecture-committee--2020.ashx

Schrage, D. P., & Lewis, W. (2020). *It is time for Army Aviation to move to a development assurance approach for including open integrated modular avionics*. Paper presented at the Vertical Flight Society's 76th Annual Forum & Technology Display, Virginia Beach, VA, 2020.

Strout, N. (2021, May 12). *The Army is making tank upgrades as simple as switching video game cartridges*. C4ISRNET. https://www.c4isrnet.com/battlefield-tech/2021/05/11/the-army-is-making-tank-upgrades-as-simple-as-switching-video-game-cartridges/

Tate, D. (2021, May 31). *NPS Acquisition Research Symposium: Panel 21 – Principles for Managing Software Acquisition*. https://www.youtube.com/watch?v=G8pXqvTlyjw&t=2s

Zimmerman, P., Ofori, M., Barrett, D., Soler, J., & Harriman, A. (2019). Considerations and examples of a modular open systems approach in defense systems. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, *16*(4), 373–388. https://doi.org/10.1177/1548512917751281

Measuring the Effects of Federal Budget Dysfunction: Impacts of Continuing Resolutions on Public Procurement

Spencer T. Brien—Assistant Professor, Naval Postgraduate School. [stbrien@nps.edu]

Korey W. Letterle—United States Marine Corps.

Paul A. Kantner—United States Marine Corps.

Abstract

This study measures the behavioral effects of continuing resolutions by examining their impact on federal procurement activities. The restrictions imposed by continuing resolutions are explored as an example of political control over a public organization. The analysis employs a dataset describing the timing of U.S. Marine Corps purchase orders for goods and services. Individual purchase orders were sampled over a four-year period (2016−2019) that endured continuing resolutions of different lengths. The analysis examines the impact of continuing resolutions on the number of purchase orders initiated, the duration of their review period, and the dollar amount per request. The results depict multiple impacts that appear to concentrate on requests for services rather than commodities. These findings help quantify the magnitude of the disruptions caused by federal budgetary dysfunction.

Introduction

In the ideal budgetary process, the U.S. Congress passes a budget appropriation act that is signed into law by the president prior to the start of the new fiscal year. This act provides federal agencies with the legal authority to obligate funds from the U.S. Treasury to pay for labor, goods, and services. Unfortunately, timely passage of the appropriation bill is the exception rather than the norm; between 2003 and 2018 the budget was passed on time in only four years. Looking even further back, Joyce (2008, p. 954) reported that in the 32-year period ending with FY2008, the budget was passed on time only four times.

When the U.S. Congress fails to pass a spending authorization prior to the start of the fiscal year, it avoids a government shutdown by enacting a continuing resolution. This provides a temporary stopgap by authorizing federal agencies to continue to operate under significant restrictions. These restrictions include a prohibition on starting any new programs that were not authorized in the previous year's budget bill as well as blocking any new multi-year procurements (Herrmann, 2017).

In addition to these formal legal restrictions, activities by federal agencies change in a variety of subtler ways when they operate under continuing resolution authority. "Agency managers may not know their budget levels from week to week——sometimes even from day to day——and this process may go on for months" (Rubin, 2007, p. 612). Uncertainty on current and future budget authority can cause managers to behave more conservatively in their spending and operational activities than they would otherwise. This may manifest as an informal hesitancy or formal directives from agency leadership to minimize spending (Herrmann, 2017).

Public procurement is one setting that may be particularly disrupted by continuing resolution status. Agencies regularly need to purchase goods and services from the private sector to complete their organizational objectives. Continuing resolutions can put formal restrictions on procurement by prohibiting new line items that were not authorized in the previous year's budget. Even if spending on goods or services is permitted, agencies may give proposed purchases additional scrutiny to ensure that they comply with legal requirements. Additionally, the risk that when full budget authority materializes it is less than expected can create further hesitation before spending. The increased administrative burden may cause some managers to simply forgo procuring resources and instead try to operate without them (Williams & Wees, 2017). Actively planning and preparing for future periods of fiscal constraint is an important part of strategic public financial management (Brien et al., 2020).

This study seeks to measure the behavioral effects of continuing resolutions on agency procurement behavior by examining a dataset describing the approval of U.S. Marine Corps purchase requests for goods and services. The data describe the time from the initial request for a good or service by the end user to the point that the request is approved by the procurement system. The individual purchase orders are spread over three years (2016–2018) that endured continuing resolutions of different lengths and an additional fourth year (2019) that had no continuing resolution for the defense portion of the federal budget. Our analysis examines several impacts of continuing resolutions on these purchase orders. The data describe the overall quantity of purchase orders, the monetary value of each individual request, and the processing time associated with each order. Weekly variations in each of these measures are compared over the four-year observation period.

The results depict significant impacts on both the count of requests initiated and the dollar amount per request. The analysis of the length of the purchase request review period was suggestive of a CR impact, but it could not be fully differentiated from seasonal effects.

Agency Responses to Continuing Resolutions

The interaction between continuing resolutions and government acquisition is one of many settings where the tension between budgetary politics and procurement implementation plays out. Thai (2001) describes how the administration of public procurement systems is guided by two types of goals that frequently come into conflict. Procurement-type goals address the efficiency, timeliness, and effectiveness of the system to acquire the goods and services required by governmental programs. Non-procurement type goals reflect the variety of ways procurement systems are used to influence other policy objectives, such as stimulating regional economies (Preuss, 2011), protecting the environment (Palmujoki et al., 2010), encouraging small businesses (Loader, 2013; Nakabayashi, 2013), or promoting social goals such as providing economic opportunities to minority or women-owned businesses (McCrudden, 2004; Myers & Chan, 1996). Developing a procurement system that tries to optimize across these various goals will require trade-offs given the limited resources and administrative capacity available to procurement professionals.

One way to frame this tension between procurement and non-procurement goals is to view it as a defining feature of the “publicness” in governmental procurement systems. Publicness theory explores the essential differences between public and private organizations and specifically examines how political influence both imposes constraints and creates opportunities for public entities (Bozeman, 1987; Pandey, 2010; Rainey 2009). Public procurement systems are different from private-sector procurement systems because of the political influences that they are subject to. These influences include the non-procurement goals that guide many of the policies and structures of acquisition systems as well as the larger structural factors arising from the political process such as continuing resolutions. Exploring agency responses to continuing resolutions is a way to understand how the realities of publicness influence administrative practice.

Dimensional publicness theory identifies four attributes that categorize forms of political and economic influence over organizations: ownership, funding, goal setting, and control (Bozeman, 1987; Fottler, 1981; Goldstein & Naor, 2005; Perry & Rainey, 1988). The last attribute, control, is of particular relevance to the study of continuing resolutions. Control refers to the influence of government-established rules and laws over an organization’s operational activities. As more of an organization’s activity is subject to and constrained by government-imposed rules, the organization exhibits a greater degree of publicness within the control dimension. Continuing resolutions add rules and restrictions to procurement activity and are a manifestation of political control over the budget.

Continuing resolutions occur when the two political parties in Congress are unable to come to agreement over budgetary legislation. The minority party or a faction within the majority may withhold support needed to achieve a sufficient number of votes. Preventing the passage of the budget is a way to exert political power to force concessions from the majority party. Generally, these concessions relate to macro issues relating to the size and composition of the overall federal budget, and the trajectory of national debt[58]. The legal and administrative ramifications of a continuing resolution, however, create micro-level constraints on operational activities, including multiple additional restrictions on public procurement.

Although continuing resolutions are associated with the control aspect of publicness, they are not a means for congress to intentionally govern procurement activity. The restrictions triggered by a continuing resolution are relatively blunt instruments. For example, a key feature of the spending restrictions in a CR is that the spending level is not explicitly specified, but instead mandated to remain under the rate of spending in the prior year. This rate is frequently established as a percentage of prior year spending that is defined in the CR legislation (GAO, 2006; Young & Gilmore, 2019). Federal agencies are able to request exemptions for specific programs to enable new program creation, but these exceptions are rarely granted by Congress. Young and Gilmore (2019) found that less than 3% of these exception requests, known as "anomalies," are ultimately granted.

Continuing resolutions also impact federal contracting by putting limits on the duration of service contracts. The federal Antideficiency Act prohibits government employees from entering into contracts or creating other forms of obligations before the funds have been appropriated under the law for that purpose (Candreva 2017; GAO, 2006; 31 U.S.C. 1341 (a)(1)(B)). This means that years with multiple successive continuing resolutions will require a series of separate contracts to maintain continuity of service. Fiscal Year 2011 had seven distinct continuing resolutions with one that lasted only three days. This leads to an increased workload to process multiple contracts for what would have been just a single contract if full budgetary authority had been granted at the beginning of the fiscal year (Bartels, 2018). Creating multiple contracts duplicates effort and increases the administrative burden of the procurement process.

The restrictions triggered by a continuing resolution create environmental conditions that influence procurement activity. Rubin (2007) found that administrators are uncertain about when full budget authority will ultimately be granted and what the final budget amount will be. This uncertainty causes administrators to act more conservatively when executing budget authority. The additional administrative tasks required by continuing resolution status, such as determining whether a given purchase can be linked to the prior year's budget authority, increase the transaction costs for a given purchase and may induce administrators to postpone procurement actions until full authority is granted (Herrmann, 2017).

This study examines how continuing resolutions impact public procurement by examining outcomes early in the acquisitions life cycle, where purchase requirements are

[58] Of course, these issues are also politicized and opposition may also be based in trying to obstruct and delay the majority party's agenda.

identified and communicated by the end user to the acquisition system. This activity occurs in both governmental and non-governmental systems, but because federal procurement is subject to the rules and environmental factors that are triggered by continuing resolutions, it is public under the control definition of publicness. The first hypothesis proposed in this analysis is that fewer purchase requests per week when the federal government is in continuing resolution status. Formal restrictions will prohibit purchase requests that cannot be associated with line items that were authorized under the previous year's budget. Additionally, administrative practices, such as agency-level directives to minimize spending, and informal behaviors, including a general reluctance to incur the administrative costs of complying with continuing resolution restrictions, will deter program office officials, on the margin, from initiating new purchase requests. The combination of these formal and informal factors is predicted to reduce the number of procurement actions started during continuing resolutions.

H1: Fewer purchase requests are initiated during continuing resolution status.

The second hypothesis predicts that continuing resolutions will increase the review and approval period prior to new purchase requests being accepted by the acquisitions system. The tasks required comply with the additional legal and administrative requirements will add to the work required to review and process new purchase requests. Additionally, procurement professionals may reject and require subsequent revisions to purchase requests at a higher rate, or they may perform stricter reviews to ensure that requests comply with regulations. These behaviors are a means of compliance with the formal requirements of continuing resolutions and a risk-reduction strategy for avoiding the consequences of violating regulations. This outcome may be viewed as an example of the "fraud/red tape" dilemma, in which organizations must trade speed and efficiency in public procurement systems to mitigate fraud or other violations of administrative law (McCue et al., 2007). Review officers may act out of an abundance of caution to avoid the legal and professional consequences that result from violating the restrictions imposed by a continuing resolution. These behaviors are predicted to increase the time from the initiation of a purchase request to its ultimate acceptance.

H2: The review and approval period will be extended during continuing resolutions.

The third hypothesis predicts that the dollar amount of purchase requests initiated during continuing resolutions will be lower because of program office requirements to minimize spending. Federal agencies can restrict spending by limiting the allocation of budget execution authority to lower-level offices within their organizations. They are also able to issue directives to minimize expenditures until final budget authorizing legislation is passed. These strategies reflect the top-down nature of cutback budgeting that occurs during periods of fiscal stress (Bozeman & Straussman, 1982).

Additionally, uncertainty regarding the final budget allotment may cause end users to reduce the size of their purchase requests. Goods may be ordered at a lower quantity of units, while service agreements may be pared down to only the most essential functions. This conservativism may help mitigate the risk that when the continuing resolution is replaced with full budget authority, the ultimate spending authorization will be lower than predicted.

H3: The average dollar amount of purchase requests will decrease during continuing resolutions.

The purpose for testing these hypotheses is to better reveal the precise effects that continuing resolutions have on procurement activity. To this point, there has been no quantitative analysis of the magnitude of the distortions that these budgetary events cause. Measuring the responses helps reveal how federal acquisitions systems are impacted by the political influences that are essential to their public character.

Background on United States Marine Corps Purchase Requests

Before discussing the empirical framework used to test the hypotheses, the following section provides some background on public procurement within the Department of Defense, and on the specific system used by the United States Marine Corps that the data used in this analysis are drawn from. Officials in the Department of Defense have recently renewed their efforts to measure and improve the speed of defense acquisitions (Berteau, 2018). Part of this improvement effort is to more clearly measure the different milestones in the life of a procurement action. The Federal Acquisition Regulations (FAR) define these milestones as beginning,

> at the point when agency needs are established and includes the description of requirements to satisfy agency needs, solicitation and selection of sources, award of contracts, contract financing, contract performance, contract administration, and those technical and management functions directly related to the process of fulfilling agency needs by contract. (FAR, 2020a, 2.101)

The time from identifying agency requirements to the time of awarding a contract for the good or service is defined as the Total Acquisition Lead Time (TALT; Kair, 1996). As depicted in Figure 1, TALT is divided into two components that are separated by the acceptance of the purchase request initiated by the end user that will ultimately use the procured resource. The first portion of the timeline is the preapproval period, defined by Letterle and Kantner (2019) as the Purchase Request Acceptance Lead Time (PRALT). This component of the acquisitions process is not systematically tracked or recorded in procurement management systems. The latter portion of the timeline that follows request acceptance is known as the Procurement Acquisition Lead Time (PALT) and is tracked for most contracting activities. PALT is the primary metric used within the Department of Defense (DoD) to evaluate the effectiveness and efficiency of contracting activities.

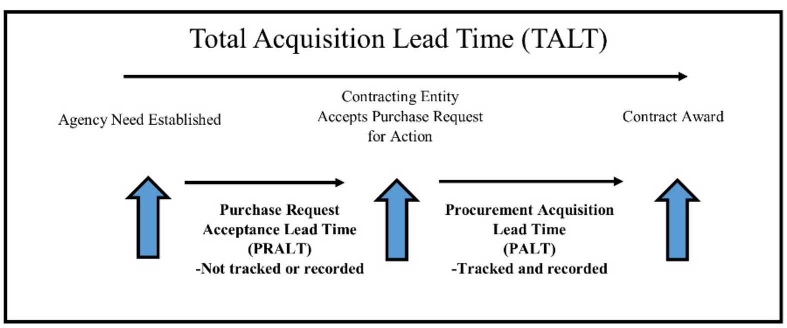


Figure 1. Purchase Request Processing Timeline (Letterle and Kantner, 2019)

The efficiency of the contracting activities that occur in the post-approval PALT phase of the acquisition process is influenced by the quality of the work done in the earlier PRALT phase. While the personnel completing the latter tasks in the PALT stage are procurement professionals, much of the early work during the PRALT phase is conducted by the end users of the procured resources, who may not be trained in acquisitions or contracting regulations. For example, Part 11 of the FAR requires end users to describe the physical characteristics of the requested resources and include them in a Statement of Work (FAR, 2020b, 11.101-11.107). If purchase requests submitted by the end user are not compliant with regulations, they are returned by the contracting professionals. Multiple rounds of revision and resubmission are not uncommon during the PRALT phase.

This lack of training and attention to the preapproval phase of public procurement has been identified in other federal agencies. The Government Accountability Office (GAO) determined that the Federal Emergency Management Agency (FEMA) was lacking in the implementation of controls over the pre-solicitation phase of contracting. Similar to the DoD, FEMA's acquisition planning process follows a two-part phase divided by the submission of an acquisition request to the contracting office. Among the GAO's recommendations were to improve the guidance for program office personnel to help implement an acquisition planning schedule (GAO, 2018).

The primary metric of contract performance used by the DoD, PALT, begins with the acceptance of a purchase request by the contracting office. All prior work by the program office during the pre-approval PRALT period, regardless of the time delay and how many revisions to the purchase request are completed, is excluded from the PALT measure. Efforts to improve the efficiency of the PRALT period must overcome the high turnover rates of military personnel within the program offices. Typically, uniformed personnel transfer occupations every two to three years. This leads to a workforce that is relatively unfamiliar with the regulatory framework surrounding defense acquisitions and is subsequently prone to making frequent errors in purchase requests that require multiple rounds of revision.

Data

The data for this analysis are obtained from the United States Marine Corps (USMC) PR Builder office, which manages the USMC's procurement record keeping system. The PR Builder office provided a randomly selected sample of purchase request records approved during the four fiscal years spanning 2016 to 2019. Individual purchase requests are identified with a Standard Documents Number (SDN). The transaction history of each purchase request, including the initial creation and submission, the final acceptance, and each intermediate determination and revision are described in the data. The records also contain details about the purchase including price and quantity information, and a descriptive field identifying whether the request is for a good or service. Additionally, the records identify the supply officer responsible for entering the purchase request into the system as well as the reviewing official that makes the determination of whether to accept or return the purchase request for revision. The names identifying individual DoD personnel were recoded with numeric identifiers and then restructured as a series of binary indicator variables identifying the supply officer and the reviewing official that processed each individual purchase request.

Defense acquisition regulations impose increased evaluation requirements for purchase requests that exceed the simplified acquisition threshold, established at $250,000. For the sake of comparison, this analysis focuses exclusively on requests that fall below that threshold. The first step in data cleaning was to identify and exclude all requests that exceeded that level. Additional records were excluded that had a total purchase value that was either set to 0 or was negative[59]. After removing other records that had missing or incomplete data, the remaining sample comprises 1,074 distinct purchase requests.

[59] The unit price and total order line amount that make up the total purchase value for individual requests in our data are entered into the system by the end user and supply officers that initially communicate purchase requests to the acquisitions system. These amounts may be subject to change during the contracting process.

The identification codes for the individual supply officers and reviewing officials control for a variety of unobserved effects. Individual performance may influence the time required to process purchase requests. Additionally, the purchase request system may route specific kinds of purchase requests to individual supply officers. For example, there are different regulations governing the procurement of commercial items versus negotiated contracts for services.[60] Different types of purchase order requests can be routed based on the expertise and the training of the supply officer for the specific kind of procurement activity. This specialization could help improve the efficiency of the purchase order review process. Controlling for the personnel involved in the review process may help identify both the content of the purchase requests and personnel efficiency effects. If present, this effect would be consistent with the findings of Decarolis et al. (2018) that explored the relationship between bureaucratic competence and procurement outcomes. A total of 28 supply officers and 13 reviewing officials are identified in the data and a corresponding number of identifying indicator variables are included in the model.

Individual purchase requests are first identified by a "date created" field. For the purposes of this analysis, this date is used as the starting point for PRALT. Ideally, PRALT would start earlier than this at the point that a need for a new resource is identified by the end user. There is no standardized process across government entities for recognizing requirements, however, and the speed with which different organizations identify their needs may also vary considerably. Given this limitation, this analysis uses the date created field as the initial point that a purchase request is made known to the acquisitions system. The final purchase request acceptance date is used to identify the end of the PRALT period and the start of PALT. Taking the difference between these two dates generates a count of the number of days that the PRALT period lasts for the purchase request.

In addition to the initiation and final acceptance dates, the purchase records include fields indicating a required date of delivery (RDD) and the start and end to a period of performance (POP). Purchase requests have either an RDD date or a POP range. Inspection of the descriptive fields indicates that the requests with RDD dates are goods, while the requests with POP ranges are services. We conducted additional manual inspection of the purchase request descriptive fields to generate an indicator variable (GOOD) that differentiates between goods and services. This indicator is then used to examine whether the responses to continuing resolutions vary for the two types of resources.

Theoretically, goods and services may fare differently under continuing resolutions if government officials are less tolerant of risk and uncertainty in their public procurement activities. Service contracts that require the development of performance-based metrics for complex activities may be perceived as higher risk (Brown et al., 2009; Martin, 2002). This perceived risk may cause service contracts to undergo additional scrutiny that extends the PRALT and PALT periods. Additionally, the transaction costs arising from recreating these complex contracts multiple times for each successive continuing resolution may induce officials to delay procurement of complex services until full budget authority is granted. Conservative managers may also reduce the size and scope of contracted services to reduce costs.

Many services, however, are associated with the fixed costs of operating government facilities. Utility contracts, facility maintenance and sanitation contracts, and other regular and

---

[60] Commercial items are covered under Part 12, while negotiated contracts are covered under Part 15 of the FAR (FAR, 2020a).

reoccurring functions would be easily associated with prior budgetary authority. Military facilities are unlikely to reduce power or water consumption during a continuing resolution. The predicted impacts of continuing resolutions on the quantity, size, and length of review period would likely have little impact on contracts for these fixed costs. To better distinguish between service types, the descriptive fields for all service purchase requests were reviewed and subsequently classified into eleven categories. Distinguishing between service types may help reveal the services that are most susceptible to politically induced controls.

Descriptive Analysis

Summary statistics of the purchase order data are depicted in Table 1. Of the 1,074 purchase order requests identified in the data, 396 are for goods, and 678 are for services. When full budget authority is present, the average PRALT pre-approval period lasts 65.2 days for services and 30.9 days for goods. For purchase requests initiated during continuing resolutions, this period lengthens to 101 days for services and 47.1 days for goods. The average total price of purchase requests initiated during full budget authority periods is $57,544 for Services and $31,618 for goods. During continuing resolutions, the average price for services drops to $39,281, while the average price for commodities rises to $45,917.

The creation of purchase orders is expected to reflect the seasonal variation in overall defense procurement. Defense spending increases towards the end of the fiscal year as managers follow the “use it or lose it” pressures to obligate funds before their budgetary authority expires (Hurley et al., 2014). These pressures encourage a surge in procurement activity in the final months and weeks of the fiscal year (Liebman & Mahoney, 2017).

Empirical efforts to measure the effect of continuing resolutions must also address the seasonal effects that simultaneously affect procurement activities (Fichtner, 2014; Hurley et al., 2014). Continuing resolutions, by definition, start at the beginning of the fiscal year and continue, with the exception of full government shutdowns, until a budget authorization bill has been enacted. Under perfect experimental conditions, it would be possible to randomly distribute periods of CR status throughout the year. Instead, the earlier parts of the year are more likely to be exposed to CR status. We control for the average seasonal effect by adding indicators for the quarter of the fiscal year in which each purchase order is initiated. Additionally, we include in the panel data from FY2019, when the defense portion of the federal budget was passed on time and there was no continuing resolution for defense spending. This provides an approximation of a control year that would exhibit the same spending seasonality, but without the added constraints of a CR. Figure 2 depicts the count of purchase orders initiated each month in each of the four years covered in the data. Additionally, this figure includes a dashed vertical line that indicates for FY2016–FY2019 when the full budget authorization bill was enacted for that year. No vertical line for FY2019 is included because its defense budget passed on time. The FY2016 appropriation bill was enacted on December 18, 2015. The FY2017 bill was enacted on May 5, 2017, and the FY2018 bill was enacted on March 23, 2018. For FY2016 and FY2017, federal spending was authorized up to the final enactment date with a series of continuing resolutions. In FY2018, federal spending experienced two relatively short funding gaps in January and February that occurred between the consecutive continuing resolutions. In each year, the count of monthly purchase orders initiated appears to increase once spending authorization transitions from continuing resolutions to current year budget authority.

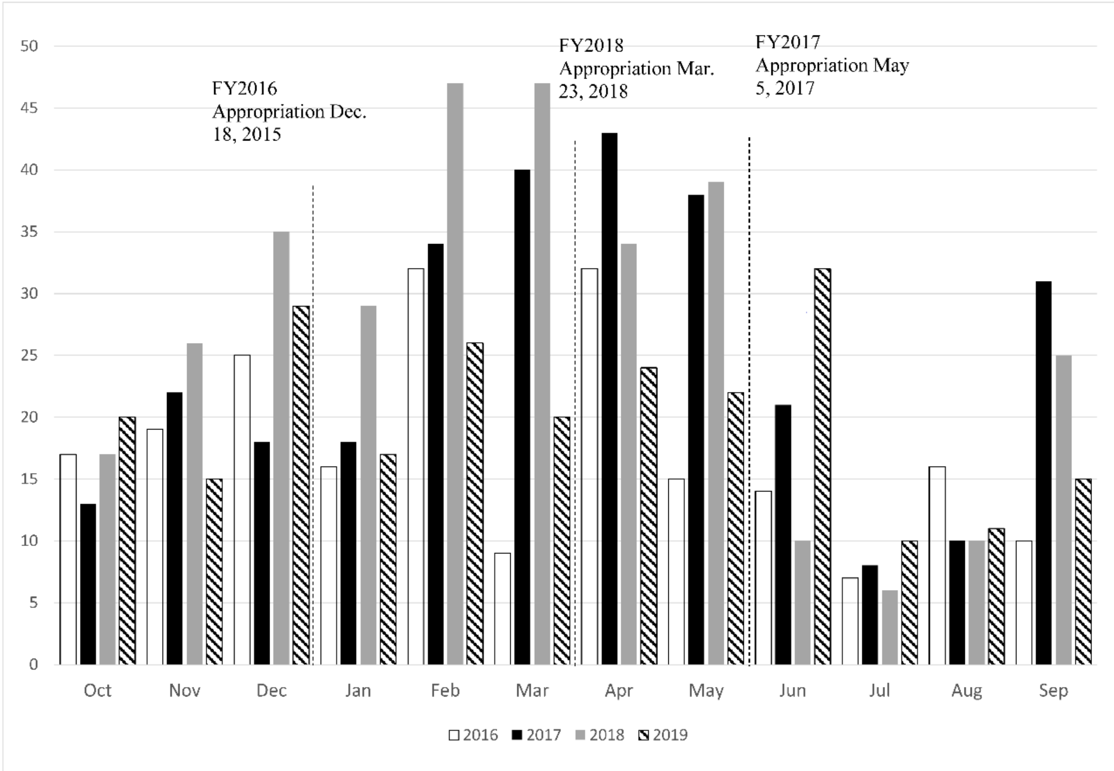


Figure 2. Count of Purchase Orders Created by Month (FY2016–2019)

Results

This section describes the quantitative analysis of three aspects of the USMC purchase orders initiated between FY2016 and FY2019. The three dependent variables examined in this study are 1) the weekly count of purchase orders initiated, 2) PRALT, or the length of time between the initial purchase order creation and its final acceptance, and 3) the total price listed in the purchase order request.

Count of Purchase Orders Initiated

The units of observation in the model of purchase order initiation are structured as weeks of the year, which generates 208 observations over the four-year period. The dependent variable in the model describes the count of the number of purchase orders initiated in a given week. Figure 3 depicts a histogram of the number of weekly purchase orders. The average number of purchase orders created per week over the four-year period was 5.16, with a variance of 15.37. Only 9 of the 208 weeks in the sample had zero purchase requests initiated and the highest count of requests in a single week was 23.

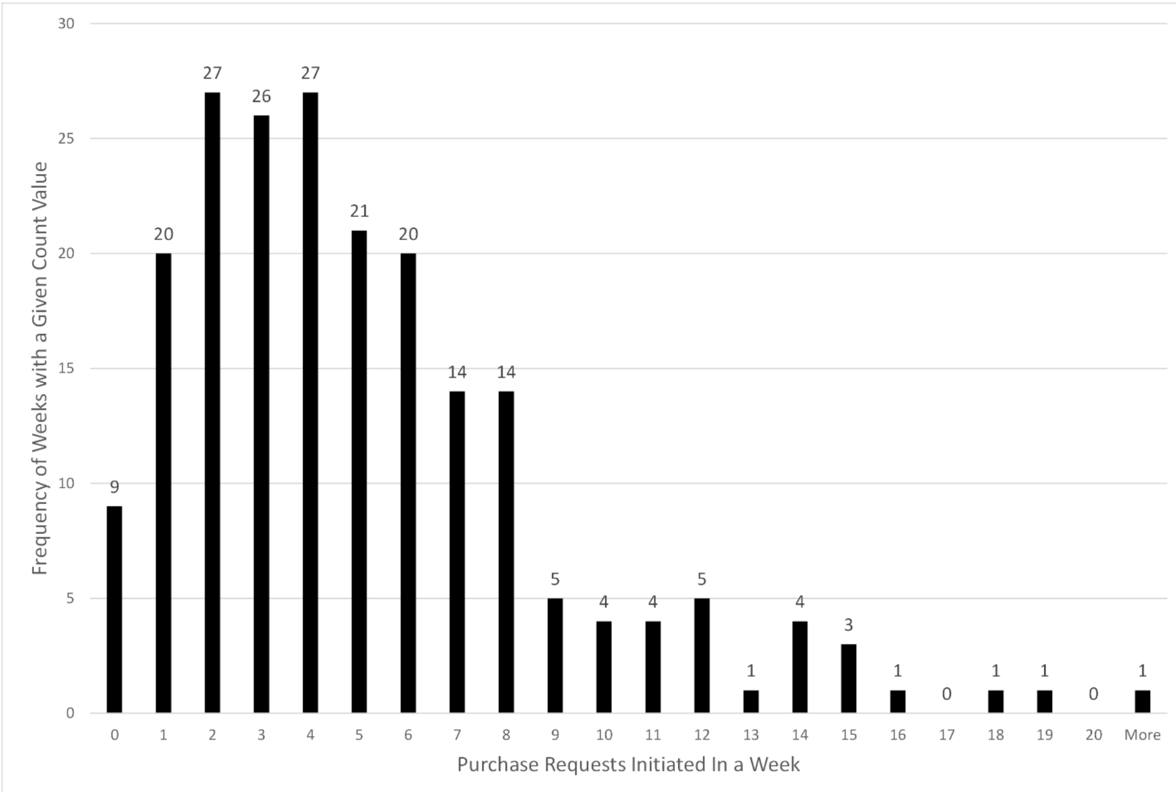


Figure 3. Histogram of the Count of Purchase Requests Initiated By Week (2016–2019)

The primary explanatory variable is an indicator that identifies whether each week occurs within that fiscal year's continuing resolution period. Control variables include indicators for the fiscal year and the quarter of the fiscal year, the latter included to control for seasonality in procurement activity. The count model is estimated for the total dataset and then for the goods and services separately. This approach depicts whether continuing resolutions have a differential effect on the number of purchase orders for goods versus services.

The model is estimated using the linear OLS estimator and the Poisson estimator. Both sets of estimates are obtained with robust standard errors. In some literatures, a Negative Binomial estimator is preferred when the variance is greater than the mean, a condition known as "over-dispersion" (Cameron & Trivedi, 2013). However, the robust Poisson estimator requires fewer restrictions[61], and is more consistent than the Negative Binomial estimator (Wooldridge 2021a, 2021b).

The estimates from both the OLS and Poisson regressions are displayed in Table 1. The estimates from the Poisson regressions have been converted to their Average Partial Effects to enable direct comparison of the marginal effects with the OLS estimates[62]. The key explanatory variable, the indicator for whether the week occurred during a continuing resolution is

[61] Wooldridge (2021b) clarifies that the Negative Binomial estimator is more efficient in the case of overdispersion only when the entire Negative Binomial distribution is correctly specified, which is a strong restriction that should not be universally assumed.
[62] The untransformed Poisson estimates are displayed in the Appendix.

statistically significant and of the predicted sign in all variants of the model. The OLS and Poisson estimates are also nearly identical. When examining total purchase requests, continuing resolution status is associated with a little more than two fewer purchase requests being initiated per week. The Column 1 OLS estimate of CR status on total orders of -2.324 is very close to the Average Partial Effect reported in Column 4 of -2.238. This is a large reduction given the average weekly number of purchase requests initiated was 5.06. The effect on the number of purchase orders initiated appears to be nearly evenly split between goods and services, though the Poisson regression indicates that service orders may be slightly more impacted by CR status than orders for goods.

It is interesting to note that the seasonal control for the first quarter of the fiscal year is statistically significant in Columns 3, 4, and 6 and marginally significant in Column 1. The combination of statistical significance in both the season controls and the CR indicator helps give assurance that the CR measure is not simply measuring the predicted seasonal pattern of government spending and the pattern of “use it or lose it.” These two factors influence procurement activity in the same downward direction, and appear to be identified separately within this analysis.

PRALT Duration Analysis

The PRALT length model is estimated using Ordinary Least Squares regression. The unit of observation for this model is the individual purchase request, yielding 1074 total records. The dependent variable is the number of days from the initial purchase request creation to its ultimate acceptance in the procurement system. The primary explanatory variable is an indicator for whether the purchase request was initiated during a continuing resolution. The control variables describe whether the request is for a good, rather than a service, the number of adjustments made to the purchase request and indicators for the fiscal year and the quarter of the fiscal year at the time of order creation. Additionally, the model includes a series of indicator variables that control for identity of the supply officers and the reviewing officials involved in processing each individual purchase request. These indicators are not displayed in the results, but can be made available upon request. The model also includes an interaction term between CR status and the goods indicator to test whether federal budget dysfunction has a differential impact on separate classes of procurement activities.

Table 3 depicts the results of the PRALT duration analysis. There are four columns in the results table. The first two columns omit the seasonal controls that are included in columns 3 and 4. Columns 2 and 4 include the interaction term between the CR indicator and the indicator that the purchase order is for a Good. In the first column, which excludes the seasonal controls, CR status is estimated to increase PRALT duration by approximately 24 days. In Column 2, which adds the interaction term, the CR coefficient estimate increases to 32.7 days, but the interaction term is negative and marginally significant. This suggests that the effect of continuing resolutions is smaller and even statistically insignificant for goods, while it remains a large effect for services. The linear combination between the CR variable and the interaction term is displayed at bottom of the table. The combination is not statistically significant, indicating the effect of the continuing resolution on lengthening PRALT times is concentrated on service procurement actions.

The results depicted in Columns 3 and 4 show that including the seasonal controls largely eliminates the effect of CR on PRALT duration. In both columns the coefficient estimate on the CR indicator is statistically insignificant. The linear combination of the CR and the interaction term’s effect is likewise insignificant. Although the estimates in Columns 3 and 4 depict effects that are in the same direction in the first two columns, they have a smaller magnitude overall. If there is an effect of CR status on PRALT length, then it is not large enough to identify it separately from the seasonality effect. The purchase order processing times for requests initiated in the first two quarters take approximately 30 days longer to complete than those initiated in the fourth quarter of the year. Those initiated in the third quarter take a little more than 16 additional days than those initiated in the fourth quarter. The inclusion of control data from 2019, which did not have a CR for defense-related spending, suggests that the variation in PRALT length is more attributable to seasonal variation in defense procurement rather than a direct impact of CR status. While the two effects appeared to be separately identified in the analysis of the count of purchase orders created, the CR effect and the seasonality effect are not separated in the PRALT length analysis

The count of adjustments made to purchase requests is significant in all models and the estimate indicates that each modification increases the length of the pre-approval period by approximately eight days. This result is not surprising, but it is interesting to quantify how iterations and adjustments to procurement actions in the pre-approval period lengthen this stage of the process.

Total Purchase Price

The purchase price model uses the same structure as the PRALT duration model. The dependent variable in this model is the real dollar amount per purchase request. Dollar amounts are normalized using the Average Annual CPI index for urban consumers. The key explanatory variable remains the indicator for whether the purchase request was initiated during a week under CR status. The other explanatory variables are also unchanged from the PRALT length model. Analyzing the total price of purchase orders revealed a modest CR effect that is mediated by whether the product is a good or a service. Table 4 depicts the OLS results from four variants of the estimated equation. Column 1 is a base model. Column 2 includes the interaction between the CR variable and the indicator that the purchase request is for a good rather than a service. Column 3 includes the seasonal controls, and Column 4 includes both seasonal controls and the interaction. The only variant where the base CR variable is statistically significant is Column 2, which indicates that purchase requests initiated during CR status have a lower price value, suggesting that orders are smaller than they would have otherwise been with full budget authority.

The coefficient on the GOOD indicator variable is negative and statistically significant in all variants of the model, suggesting that the purchase value of requests for commodities average several thousand dollars less than requests for services. The interaction term between the CR indicator and GOOD, however, is positive and statistically significant. The opposing estimates of the interaction term and the base CR indicator cause the linear combination of the two variables to be statistically insignificant, with a p-value of 0.125. This indicates that the dollar amount of purchase requests widens between goods and services during continuing resolutions. The negative value of the base CR variable only holds for services once the interaction terms are fully taken into account.

In Column 4, which includes the seasonal controls, the CR indicator is no longer statistically significant, but the interaction between the CR variable and the GOOD variable remains significant. The linear combination of the two in Column 4 is also significant, with a p-value of 0.043. This is notable because, even controlling for the seasonality effect, the differential in ordering value between goods and services associated with continuing resolutions remains.

It makes sense that services would experience a differential impact under continuing resolutions. One of the important restrictions on defense procurement triggered by a CR is that service contracts may only last for the duration of budget authority. In recent history, Congress has produced multiple consecutive continuing resolutions within a single budget year prior to the enactment of full budget authority. Procurement officers may seek to minimize or simplify service requests from end users during this period in order to reduce anticipated replication of contracting work for each subsequent CR.

One potential criticism of this analysis is that smaller value purchase requests may be less likely to be impacted by continuing resolutions. As a robustness test, we reran the analysis in Column 4 of Table 4, but excluded all purchase requests with a value less than $10,000. This restriction dropped 602 of the 1,074 observations from the estimation. The results were comfortingly similar to the full analysis. The pattern of statistical significance in the CR control, the commodity indicator (GOOD), and the interaction term between GOOD and CR was unchanged and the magnitude of the coefficients on both the interaction term and the linear combination of CR and the interaction term increased. This test confirms that the main results hold for the purchase requests most likely to be influenced by the CR restrictions.

Discussion

The above analysis examined the impact of continuing resolutions on three different aspects purchase request development. The strongest effects were observed on the count of requests initiated per week and the total dollar amount per purchase request. The regressions on the count data revealed relatively similar effects on both goods and services, though the Poisson regressions suggested a marginally larger reduction on the number of service requests initiated per week during CR status. The count regressions also produced estimates of the CR effect that were the most clearly differentiated from seasonality effects that are inherent to federal budget execution.

The estimates of CR impact on the dollar amounts per purchase request also showed significant impacts that persisted after controlling for seasonality. The results showed a widening of the differential between goods and services during CR status that suggests that the dollar amount of service requests is suppressed relative to requests for commodities when the government is operating without full budget authority. The results from analyzing PRALT length were initially suggestive of a strong CR impact, but the statistical significance of that set of findings diminished after controlling for seasonality.

What do these findings indicate for the study of publicness and how political control over the budget impacts agency procurement and administration? It is important to recognize that the political decision to enter into a continuing resolution is not an attempt to intentionally exert control over procurement behavior. The concept of the control dimension of publicness theory may need to be adapted to differentiate between intentional and unintentional control. Unintentional control would encompass the legal and administrative regulations that are triggered by political action. These restraints may not be part of the explicit goal of high level policy action, but organizations that suffer resource restrictions or other administrative burdens because of the resulting policy outcomes are experiencing the consequences of publicness.

This may be better understood by applying Moulton's framework for understanding the components of publicness (2009). The public value dominating continuing resolution policy is that elected officials want to avoid a full government shutdown in the event that a budgetary compromise has not been achieved. Any interim spending by federal agencies, however, must be controlled so that the executive branch does not usurp power from Congress over the determination of the budget. The restrictions and mechanism of continuing resolutions maintain spending allocations that Congress had previously authorized until a new budget is enacted. The conservative responses to executing budget authority at the agency level, however, are expected given the increased administrative burden continuing resolutions impose. They are natural responses to the risk and uncertainty Rubin (2007) identified that pervades financial management when full budget authority is absent. The realization of publicness is manifested in the worsening performance of organizations that are impacted by these restraints.

Organizations outside of the federal government may also experience impacts of this form of unintentional control. Certainly federal contractors are impacted by federal budgetary instability. Additionally, restrictions caused by continuing resolutions on the release of federal grants to non-profit organizations that are engaged in research, health, and other public enterprises would also be a manifestation of this form of publicness. The degree of disruption to organizational health and operations cause by continuing resolutions is another way to understand its status as a public entity.

Limitations

The implications for this analysis for agencies outside of the USMC are subject to a few limitations. First, the period of time observed in this study is affected by the Budget Control Act caps on defense spending, commonly known as sequestration. The year fixed-effects included in the model can control for some of the average impact of the reduced spending in those years. Amendments to the Budget Control Act raised the defense spending caps and largely restored the level of spending in 2018 and 2019 to the pre-sequestration trajectory, so the effect in those years would be smaller than the effect in 2016 and 2017 (McGarry, 2019). Repeating the analysis over a more extended time period that surpasses the BCA limitations would provide additional information regarding the relationship between CR status and procurement activity. As an added benefit, a longer period of analysis would also increase the variation in continuing resolution status across the observed years.

A second limitation of this study is that the detail in the data describing the actual services provided is relatively limited. The categorization of individual purchase orders as goods or services is a useful distinction, but more controls for different types of goods and services may help increase the explanatory power of the model. For example, it would be helpful to be able to differentiate utilities from other service contracts that are more tied to operational activities rather than the fixed costs of operating federal facilities. For both goods and services, controls that help separate out the complexity of the product requested may help explain variation in the length of time required to process requests prior to acceptance.

A third limitation of this paper is that the purchase orders identified in the data were all orders that were ultimately accepted for procurement. This study does not have data covering requests that were rejected and then never resubmitted. Omitting these records means that these results may underestimate the full impact of continuing resolutions on federal procurement activity.

Conclusion

Irene Rubin concludes her assessment of the “Great Unravelling” of federal budgeting norms with the lament that “it is not so much that we do not know what reforms are likely to work, but that we do not know how to motivate those who benefit from the status quo to adopt and implement the necessary reforms” (2007, p. 615). It would be grossly naïve to assume that this study will provide that motivation, but efforts to quantify the administrative burden of continuing resolutions may help influence budgetary deliberations. Providing lawmakers, agency officials, and congressional staff with evidence of the erosion of agency performance caused by the lack of full budget authority may help shift the calculus of using budget delays as a political tool.

Consider how private businesses would perform if they spent the first quarter of every year in stasis. What would be the outcome if no new products were ever introduced and no efforts to modernize were implemented during this period? What if firms made no responses to changes in market conditions during the first quarter? How would competitors react if they knew this behavior was repeated year after year? This is the current state of the federal government.

This analysis explores a narrow slice of behavior in the DoD, but other federal agencies fall under the same restrictions. National security is affected, but also policy towards education, housing and urban development, and national health. The impact of reduced agency performance under continuing resolutions is felt across society.

There are many anecdotes, news stories, and personal interviews describing the impact of continuing resolutions. This study, to our knowledge, is the first to quantify the impact of these restrictions on agency behavior and the first to explore the interaction between continuing resolutions and public procurement. Further studies that explore the behavioral responses to the uncertainties created by continuing resolutions and to the heightened regulatory framework are needed to give a more complete accounting of the costs of Congress's failing to enact a budget bill on time.

Table 1. Summary Statistics of PRALT Length and Total price, Differentiated By Good/Service and Continuing Resolution Status

		Continuing Resolution		**Full Budget Authority**		**Overall**
		Service	Good	Service	Good	
Number of Purchase Orders		148	80	530	316	Total 1,074 Goods 396 Services 678
PRALT Length in Days	Mean	101	47.1	65.2	30.9	58.7
	Std Dev	(173.5)	(76.3)	(119.1)	(50.1)	(113)
Total Price	Mean	$39,281	$45,917	$57,544	$31,618	$46,533
	Std Dev	($53,990.7)	($75,002.4)	($63,626.5)	($50,148.2)	($60,665)
Adjustments	Mean	4.8	3.3	5.1	3.3	4.4
	Std Dev	(3.8)	(3.7)	(5.4)	(3.9)	(4.7)

Table 2. Results of OLS and Poisson Regression Analyses of Continuing Resolution Status on Purchase Requests Initiated Per Week

	OLS Regressions			Average Partial Effects from Poisson Regression		
	(1)	(2)	(3)	(4)	(5)	(6)
Count of Purchase Orders	Total Orders	Service Orders	Goods Orders	Total Orders	Service Orders	Goods Orders
Indicator of Continuing Resolution (CR)	-2.324***	-1.200**	-1.125**	-2.238***	-1.237**	-0.952**
	(0.793)	(0.545)	(0.449)	(0.781)	(0.539)	(0.456)
Fiscal Year 2017	2.465***	1.400**	1.065**	2.447***	1.409**	1.024**
	(0.780)	(0.569)	(0.427)	(0.776)	(0.562)	(0.445)
Fiscal Year 2018	2.620***	1.365**	1.255***	2.542***	1.328**	1.207***
	(0.766)	(0.569)	(0.431)	(0.722)	(0.538)	(0.415)
Fiscal Year 2019	0.066	0.535	-0.469	0.299	0.593	-0.278
	(0.621)	(0.471)	(0.328)	(0.530)	(0.417)	(0.276)
FYQ1 (Oct.−Dec.)	-1.577*	-0.589	-0.988**	-2.143**	-0.792	-1.452**
	(0.823)	(0.535)	(0.478)	(0.903)	(0.569)	(0.574)
FYQ2 (Jan.−Mar.)	-0.311	-0.112	-0.200	-0.172	-0.0625	-0.103
	(0.759)	(0.488)	(0.471)	(0.706)	(0.471)	(0.427)
FYQ3 (Apr.−June)	0.390	0.708	-0.317	0.345	0.607	-0.244
	(0.838)	(0.620)	(0.449)	(0.667)	(0.501)	(0.353)
Constant	4.943***	2.790***	2.153***			
	(0.622)	(0.443)	(0.384)			
Observations	208	208	208	208	208	208
R-Squared	0.199	0.118	0.183			
Psedudo R-Squared from Poisson				0.102	0.0580	0.105

Note: Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3. OLS Regression Estimates of the Time from Initial Purchase Request Submission to Acceptance in the PR Builder System

	(1)	(2)	(3)	(4)
Days to Purchase Request Acceptance	Base Model	Goods Interaction	Seasonal Controls	Seasonal Controls and Goods Interaction
Indicator of Continuing Resolution (CR)	23.918**	32.741**	5.617	15.416
	(10.170)	(14.389)	(11.749)	(15.890)
Purchase Order for Goods (GOOD)	-9.736*	-4.508	-9.349*	-3.666
	(4.964)	(5.643)	(4.890)	(5.462)
Interaction CR*GOOD		-24.831*		-27.165*
		(14.966)		(15.542)
Count of Adjustments Made to the PR	8.390***	8.386***	8.365***	8.366***
	(1.136)	(1.139)	(1.138)	(1.142)
Fiscal Year 2017	19.121	20.292	25.497**	26.791**
	(12.411)	(12.338)	(12.644)	(12.586)
Fiscal Year 2018	26.975**	27.291**	28.350***	28.645***
	(10.826)	(10.768)	(10.654)	(10.603)
Fiscal Year 2019	2.367	3.410	-1.530	-0.383
	(9.584)	(9.539)	(9.281)	(9.241)
FYQ1 (Oct.−Dec.)			29.823**	29.330**
			(12.026)	(12.109)
FYQ2 (Jan.−Mar.)			30.780***	31.654***
			(7.720)	(7.843)
FYQ3 (Apr.−June)			16.113**	16.489**
			(7.644)	(7.616)
Constant	-27.712***	-30.105***	-42.415***	-45.358***
	(10.492)	(10.384)	(12.505)	(12.378)
Observations	1,074	1,074	1,074	1,074
R-squared	0.275	0.277	0.282	0.285
Linear combination of CR and CR*GOOD		7.909		-11.749
		(7.689)		(9.360)

Robust standard errors in parentheses

Note: A series of dummy variables controlling for the unique identities of the supply officers and reviewing officials involved in processing the purchase requests were also included in the model; 27 dummies for the supply officers and 12 dummies for the reviewing officials were included. The estimates of these control variables are not included in the table, but available upon request.

*** p<0.01, ** p<0.05, * p<0.1

Table 4. OLS Regression Estimates of the Real Dollar Amount for PR Builder Purchase Requests

Real Dollar Amount Per Purchase Request	(1) Base Model	(2) Goods Interaction	(3) Seasonal Controls	(4) Seasonal Controls and Goods Interaction
Indicator of Continuing Resolution (CR)	-2,238.1	-6,712.8***	557.9	-4,016.2
	(2,023.7)	(2,187.5)	(2,825.2)	(2,952.4)
Purchase Order for Goods (GOOD)	-5,189.9***	-7,841.5***	-5,082.0***	-7,734.7***
	(1,641.5)	(1,822.1)	(1,637.9)	(1,821.3)
Interaction CR*GOOD		12,594.1***		12,680.4***
		(4,336.1)		(4,363.9)
Count of Adjustments Made to PR	444.5**	446.6**	461.4**	460.9**
	(213.5)	(211.2)	(214.6)	(211.9)
Fiscal Year 2017	4,471.7	3,877.8	3,574.1	2,970.1
	(2,748.9)	(2,714.8)	(2,822.3)	(2,793.1)
Fiscal Year 2018	1,257.8	1,097.6	821.2	683.4
	(2,546.7)	(2,528.8)	(2,560.6)	(2,540.9)
Fiscal Year 2019	2,462.4	1,933.4	3,049.5	2,514.5
	(2,848.4)	(2,826.2)	(2,881.5)	(2,860.8)
FYQ1 (Oct.−Dec.)			-2,702.8	-2,472.5
			(3,310.3)	(3,275.7)
FYQ2 (Jan.−Mar.)			-2,171.7	-2,579.5
			(2,297.3)	(2,306.2)
FYQ3 (Apr.−June)			2,361.8	2,186.0
			(1,982.1)	(1,977.4)
Constant	16,061.1***	17,274.5***	14,952.9***	16,326.4***
	(3,931.5)	(3,940.9)	(4,158.9)	(4,172.6)
Observations	1,074	1,074	1,074	1,074
R-squared	0.096	0.105	0.010	0.109
Linear combination of CR and CR*GOOD		5881.2		8664.3**
		(3826.8)		(4285.2)

Robust standard errors in parentheses

Dollar amounts normalized using Annual Average CPI index for urban consumers.

Note: A series of dummy variables controlling for the unique identities of the supply officers and reviewing officials involved in processing the purchase requests were also included in the model; 27 dummies for the supply officers and 12 dummies for the reviewing officials were included. The estimates of these control variables are not included in the table, but available upon request.

*** p<0.01, ** p<0.05, * p<0.1

References

Bartels, F. (2018). *Continuing resolutions invariably harm national defense*. The Heritage Foundation. https://www.heritage.org/defense/report/continuing-resolutions-invariably-harm-national-defense

Berteau, D. (2018). DoD can reduce time to contract. *The Federal Times,* 30.

Bozeman, B. (1987). *All organizations are public: Bridging public and private organizational theories*. Jossey-Bass Inc Pub.

Brien, S. T., Candreva, P. J., Hansen, S. C., & Pizzini, M. 2020. Looking forward to cuts: A simulation decision tool for cutback management. *Public Budgeting & Finance, 40*(1) 3–21.

Brown, T. L., Potoski, M., & Van Slyke, D. M. (2009). Contracting for complex products. *Journal of Public Administration Research and Theory, 20*(Supplement 1), i41–58.

Cameron, A. C., & Trivedi, P. K. (2001). Essentials of count data regression. *A Companion to Theoretical Econometrics*.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data*. Cambridge University Press.

Decarolis, F., Giuffrida, L. M., Iossa, E., Mollisi, V., & Spagnolo, G. (2018). *Bureaucratic competence and procurement outcomes* [Working Paper]. National Bureau of Economic Research. http://www.nber.org/papers/w24201

FAR. (2020a). https://www.acquisition.gov/content/part-2-definitions-words-and-terms.

FAR. (2020b). 11. https://www.acquisition.gov/content/part-2-definitions-words-and-terms.

Fichtner, J. (2014). Curbing the splurge in year-end federal government spending: Reforming "use Iit or lose it policies [Mercatus working paper]. https://ssrn.com/abstract=3191368

GAO. (2018). *2017 disaster contracting: Action needed to better ensure more effective use and management of advance contracts*. https://www.gao.gov/products/GAO-19-93

Herrmann, R. M. (2017). The pitfalls of new starts during a continuing resolution. *Army Lawyer, 2017*(8), 5–13.

Hurley, W. J., Brimberg, J., & Fisher, B. (2014). Use it or lose it: On the incentives to spend annual defence operating budgets. *Defence and Peace Economics, 25*(4), 401–413.

Joyce, P. G. (2008). Does more (or even better) information lead to better budgeting? A new perspective. *Journal of Policy Analysis and Management, 27*(4), 945–960.

Kair, L. (1996). *An analysis of total acquisition lead time for small purchase actions in an operational contracting environment.* Air Force Institute of Technology.

Letterle, K. W., & Kantner, P. A. (2019). *An analysis of contracting activity purchase request acceptance lead time for USMC using unit acquisitions under the simplified acquisitions threshold.* Naval Postgraduate School.

Liebman, J. B., & Mahoney, N. (2017). Do expiring budgets lead to wasteful year-end spending? Evidence from federal procurement. *American Economic Review, 107*(11), 3510–3549.

Loader, K. (2013). Is public procurement a successful small business support policy? A review of the evidence. *Environment and Planning C: Government and Policy, 31*(1), 39–55.

Martin, L. L. (2002). Performance-based contracting for human services: Lessons for public procurement. *Journal of Public Procurement, 2*(1), 55–71.

McCrudden, C. (2004). Using public procurement to achieve social outcomes. *Natural Resources Forum, 28*(4), 257–267.

McCue, C. P., Buffington, K. W., & Howell, A. D. (2007). The fraud/red tape dilemma in public procurement. In *Public Procurement: International Cases and Commentary*, Routledge, 247–263.

McGarry, B. W. (2018). *The defense budget and the budget control act: Frequently asked questions* (44039). Congressional Research Service.

Moulton, S. (2009). Putting together the publicness puzzle: A framework for realized publicness. *Public Administration Review*, *69.5*(2009), 889–900.

Myers, S. L., & Chan, T. (1996). Who benefits from minority business set-asides? The case of New Jersey. *Journal of Policy Analysis and Management, 15*(2), 202–226.

Nakabayashi, J. (2013). Small business set-asides in procurement auctions: An empirical analysis. *Journal of Public Economics, 100*, 28–44.

Palmujoki, A., Parikka-Alhola, K., & Ekroos, A. (2010). Green public procurement: Analysis on the use of environmental criteria in contracts. *Review of European Community & International Environmental Law, 19*(2), 250–262.

Pandey, S. K. (2010). Cutback management and the paradox of publicness. *Public Administration Review, 70*(4), 564–571.
Preuss, L. (2011). On the contribution of public procurement to entrepreneurship and small business policy. *Entrepreneurship & Regional Development, 23*(9–10), 787–814.
Rainey, H. G. (2009). *Understanding and managing public organizations*. John Wiley & Sons.
Rubin, I. (2007). The great unraveling: Federal budgeting, 1998–2006. *Public Administration Review, 67*(4), 608–617.
Thai, K. V. (2001). Public procurement re-examined. *Journal of Public Procurement, 1*(1), 9–50.
Williams, L. M., & Wees, D. P. (2016). *FY2017 defense spending under an interim continuing resolution (CR): In brief*. Congressional Research Service.
Woolridge, J. [@jmwoolridge]. (2021, April 15). *Confusion about the choice between Baysian and classical methods* [Tweet]. Twitter. https://twitter.com/jmwoolridge/status/1387747039667855390
Woolridge, J. [@jmwoolridge]. (2021, July 17). *"Efficiency" is another world that we often use too loosely* [Tweet]. Twitter. https://twitter.com/jmwoolridge/status/1416420329342345219

Appendix

Poisson Regression of Purchase Orders Initiated Per Week

	Poisson Regressions		
	(1)	(2)	(3)
Count of Purchase Orders	Total Orders	Service Orders	Goods Orders
Indicator of Continuing Resolution	-0.433***	-0.380**	-0.500**
	(0.147)	(0.161)	(0.234)
Fiscal Year 2017	0.485***	0.453***	0.530**
	(0.143)	(0.173)	(0.216)
Fiscal Year 2018	0.500***	0.432**	0.601***
	(0.134)	(0.168)	(0.197)
Fiscal Year 2019	0.074	0.216	-0.211
	(0.130)	(0.152)	(0.211)
FYQ1 (Oct.−Dec.)	-0.415**	-0.243	-0.762**
	(0.174)	(0.175)	(0.297)
FYQ2 (Jan.−Mar.)	-0.033	-0.019	-0.054
	(0.137)	(0.144)	(0.224)
FYQ3 (Apr.−June)	0.067	0.186	-0.128
	(0.129)	(0.152)	(0.184)
Constant	1.539***	0.993***	0.674***
	(0.119)	(0.144)	(0.191)
Observations	208	208	208
LL	-557.8	-472.7	-385.7
PseudoR2	0.102	0.0580	0.105

Note: Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

PANEL 25. DIGITAL ENGINEERING ACROSS THE ACQUISITION LIFECYCLE

Thursday, May 12, 2022	
1:55 p.m. – 3:10 p.m..	**Chair: Gaurang R. Dävé,** Cyber Technology Officer, Marine Corps Systems Command ***The Key Elements and Drivers of the Defense Acquisition System*** Camryn Burley, Stevens Institute of Technology Philip S. Antón, Stevens Institute of Technology John Boardman, Stevens Institute of Technology Dinesh Verma, Stevens Institute of Technology ***Determining a Digital Engineering Framework: A Systematic Review of What and How to Digitalize*** Stephen Waugh, Johns Hopkins University Applied Physics Laboratory ***Instrumenting the Acquisition Design Process: Developing Methods for Engineering Process Metrics Capture and Analysis*** Lorri Bennett, Penn State University Jay Martin, Penn State University Simon Miller, Penn State University Michael Yukish, Penn State University

Gaurang R. Dävé— serves as Cyber Technology Officer for the Marine Corps Systems Command (MCSC). In this role, he serves as the principal cyber advisor and technical expert to the Commander, Executive Director and Chief Engineer, MCSC; and Program Executive Office Land Systems (PEOLS) and subordinate activities, for Cyberspace, Cybersecurity, and information technology (IT).

Mr. Dävé previously served as Cyber Advisor at MCSC. In this role, Mr. Dävé provided technical cyber Subject Matter Expertise (SME) for cyber policy, budgeting, evaluating new technologies, representing command senior leadership at DoD, DoN, and within USMC for cyber related matters.

Prior to joining MCSC, Mr. Dävé served as Senior Cyber Technical Advisor and Cybersecurity Program Director for Naval Surface Warfare Center (NSWC) Dahlgren. In this capacity, Mr. Dävé led the Navy's combat systems cybersecurity portfolio across the Warfare Systems Program Office. Mr. Dävé provided technical leadership, guidance, and program management leadership. Prior to that, Mr. Dävé served as Cybersecurity Risk Posture lead for the Office of the Chief of Naval Operations, OPNAV N2N6 Cybersecurity Division at the Pentagon. His contributions included synchronizing cyber strategy, standards and requirements, evaluating and prioritizing investments, providing oversight, and resource sponsorship.

Mr. Dävé began his career as a software engineer and served in technical, program management, and supervisory positions across the joint services. Mr. Dävé led the Chemical, Biological, Radiological (CBR) Analysis, Testing, and Systems Engineering Branch at NSWC, Dahlgren Division. Under his leadership, the branch provided SME in CBR modeling & simulation, threat analysis, executed White House initiatives for Medical Countermeasures, and led several ACAT-3 efforts. Mr. Dävé served as CBR Thrust Area Manager for Major Defense Acquisition Program at Joint Science & Technology Office led by US Army. Prior to joining government service, Mr. Dävé worked in private industry at International Business Machines (IBM). As a Security Software Engineer, Mr. Dävé led several software security program development efforts designed to protect the company's network.

Mr. Dävé holds Bachelor of Science degrees in Computer Science and Biochemical Pharmacology from State University of New York at Buffalo. Master of Science degree in Systems Engineering with focus in Operational Research from George Mason University. Graduate degree in Cyber Engineering from University of Maryland. He is DAWIA Level III certified in Systems Engineering.

The Key Elements and Drivers of the Defense Acquisition System

Camryn Burley—received her Bachelor of Science in Systems Engineering from the University of Virginia, and then completed this work as her capstone experience for her master's in Space Systems Engineering from Stevens Institute of Technology. After graduating from Stevens in December 2021, she now works as a Systems Engineer at L3Harris Technologies in Clifton, NJ. [cburley@stevens.edu]

Philip S. Antón—conducts research on acquisition, policy, AI, cybersecurity, emerging technologies, and test infrastructure. He is the Chief Scientist, Acquisition Innovation and Research Center, at the Stevens Institute of Technology. Prior to this, he was a Senior Information Scientist and Director at RAND. He served two Pentagon tours as the Deputy Director for Acquisition Policy Analysis, receiving the Secretary of Defense Medal for Outstanding Public Service in 2017. From 1992 to 1998, Antón managed and conducted artificial intelligence research at MITRE. Antón earned his PhD and MS in information and computer science from UCI (computational neuroscience and artificial intelligence) and his BS in computer engineering from UCLA. [panton@stevens.edu]

John Boardman—received the PhD (1970) and the BEng (1967) in Electrical Engineering from the University of Liverpool. He held tenured positions at four British universities including Portsmouth where he was founding head of the School of Systems Engineering and GEC Marconi Professor of Systems Engineering, and later Dean of the Faculty of Technology. He was Distinguished Service Professor (2005–2012) at Stevens Institute of Technology where he helped establish the School of Systems and Enterprises. He is a Fellow of IET and INCOSE. His area of research is the application of systemic thinking to enterprise architecting. [johntboardman@hotmail.co.uk]

Dinesh Verma—is a Professor of Systems Engineering at Stevens Institute of Technology. He received the PhD (1994) and the MS (1991) in Industrial and Systems Engineering from Virginia Tech. He served as the Founding Dean of the School of Systems and Enterprises at Stevens for 10 years, from 2007 through 2016. He currently serves as the Executive Director of the Systems Engineering Research Center (SERC), a U.S. Department of Defense University Affiliated Research Center (UARC) focused on systems engineering research, and the Acquisition Innovation Research Center (AIRC). During his 20 years at Stevens, he has successfully proposed research and academic programs exceeding $175 million in value. Prior to this role, he served as Technical Director at Lockheed Martin Undersea Systems, in Manassas, VA, in the area of adapted systems and supportability engineering. [dverma@stevens.edu]

Abstract

A systemic diagram (systemigram) was developed to provide a systems view of the key elements and drivers of the complex defense acquisition system in the United States. An iterative process was used to develop the systemigram, after assessing the basic relationships among key actors and organizations within the system. The diagram provides a high-level overview of the Department of Defense ecosystem as it relates to acquisition, addressing the lack of available high-level visual representations of the overall acquisition system elements and their basic interactions within the literature. Using this diagram, individuals unfamiliar with the defense acquisition system can become better acquainted with it, while those familiar with defense acquisition are provided with a useful artifact to stimulate shared understanding, spark conversations about how to improve acquisition outcomes, and focus on the key inputs, processes, and ultimate goal of military capability.

Keywords: systemigram, acquisition innovation, acquisition overview

Introduction

Natural language (i.e., prose) descriptions of complex systems, such as the defense acquisition system, can be more effectively conveyed visually through the use of systemic diagrams (systemigrams; Mehler et al., 2010). Systemigrams can represent an entire document's worth of information on a single page, which can be more easily and quickly consumed by an audience. The audience can then use the systemigram as a means to converge on a shared mental model of the system and to elicit insight through conversations about the diagram (Mehler et al., 2010). Systemigrams tell a story about the given system and are presented as a series of scenes to communicate the message of the diagram itself, as well as the message of the system (Blair et al., 2007). Each scene is comprised of a subnet of the diagram (Blair et al., 2007), displaying or adding only a small number of nodes (noun phrases) and arcs (verb phrases) at a time. The main purpose or takeaway of the system is represented as a sentence along the diagonal of the diagram and is called the mainstay (Sauser, 2019). The last node of the mainstay is the overall goal or objective of the system (Sauser, 2019).

A systemigram was chosen to provide a systems view of the Department of Defense (DoD) ecosystem as it relates to defense acquisition. The systemigram developed for this paper will be referred to as the Defense Acquisition Systemigram. The high-level Defense Acquisition Systemigram is especially useful for an audience of individuals just becoming acquainted with defense acquisition, such as professors beginning involvement in defense acquisition research or new members of Congress. The Defense Acquisition Systemigram may also have applications for those familiar with defense acquisition in providing context, recentering them on the ultimate goal of defense acquisition, and providing a means to have conversations about acquisition innovation.

Related Work

The Defense Acquisition Systemigram detailed in this paper is unique in its presentation of the defense acquisition process at a high level. Though at least two other systemigrams regarding defense acquisition exist, they are either constructed for a different audience, including finer details, or focus on the conditions that lead to the success or failure of acquisition projects, instead of the functions of acquisition. Outside of systemigrams, it is difficult to locate another visual depiction of the overall defense acquisition system. The Defense Acquisition Systemigram helps to fill a gap in the literature concerning high-level, visual depictions of the defense acquisition system.

Figure 1 displays the first of the two systemigrams about defense acquisition in the literature. The systemigram in Figure 1 aims to visualize the “structure, function, and process of the defense acquisition system and the extended acquisition enterprise” (Cilli et al., 2016). Figure 1 was developed as an interpretation of the, at the time, latest revision to the defense acquisition system and displays one pathway through the process. While useful, Figure 1 has been created for a different audience than the Defense Acquisition Systemigram as it includes more detail and utilizes terms related to the steps in the processes involved, rather than references to stakeholders, communities, and their relationships that provide key elements needed to acquire and field military capabilities.

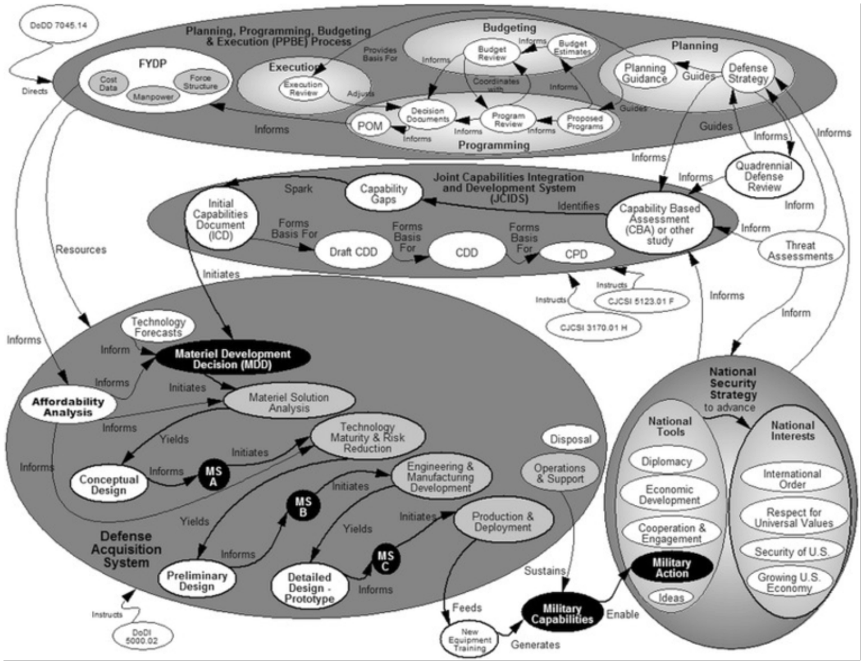


Figure 1. Systemigram from Cilli et al. (2016)

The second systemigram previously created about the defense acquisition system is reproduced in Figure 2. The focus of this systemigram is the conditions leading to a successful or unsuccessful acquisition project, which is fundamentally different from the Defense Acquisition Systemigram's concentration on the process as a whole. Figure 2 focuses on the actors of combat veterans, engineers, Congress, contractors, and program managers. The scenarios to be avoided are program delays or cancellation or increases in overall program cost or unit cost. It poses that the ultimate goal of the defense acquisition system is successful conflict outcome.

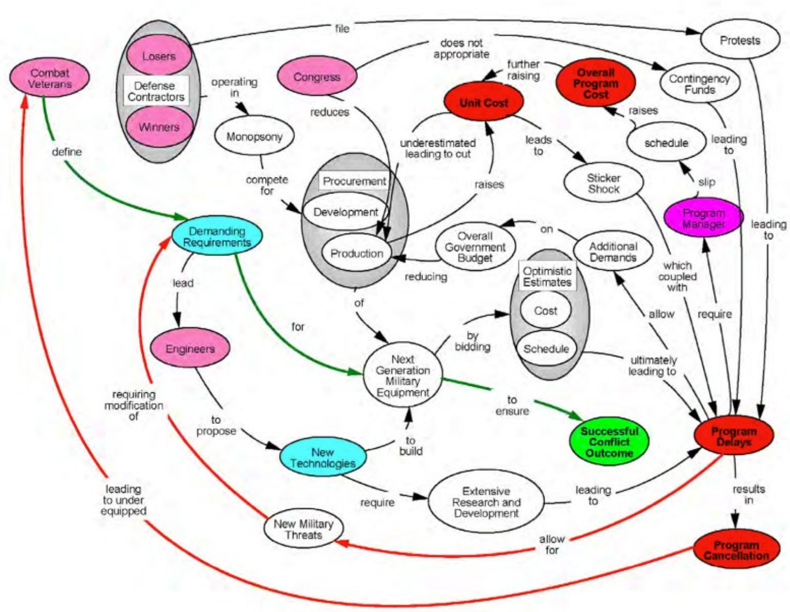


Figure 2. Systemigram from Wade & Batra (2019)

Process

Before beginning the Defense Acquisition Systemigram, a list was generated of actors involved in the system and the functions that they perform. Then, research was conducted on these actors, which are departments and organizations within and relating to the DoD. A hierarchy of these departments and organizations was developed to understand the basic relationships existing among actors and functions. An iterative process followed the creation of an initial draft of the systemigram. Drafts were presented to a subject-matter expert for feedback (P. S. Anton, personal communication, April–October 2021) and later improved. The aim of iterating was to arrive at the most accurate, useful, and visually understandable diagram form of the narrative. Important to this project was correctly stating the mainstay for the system and deciding on the level of detail for the systemigram, including key inputs, sources, and feedback loops.

Defense Acquisition Systemigram

Overview

The final Defense Acquisition Systemigram presents a high-level overview of the defense acquisition system, including how the Acquisition Community interfaces with other functions and organizations outside of itself. The diagram in its entirety is shown in Figure 3 and will be constructed, scene by scene, in the following sections.

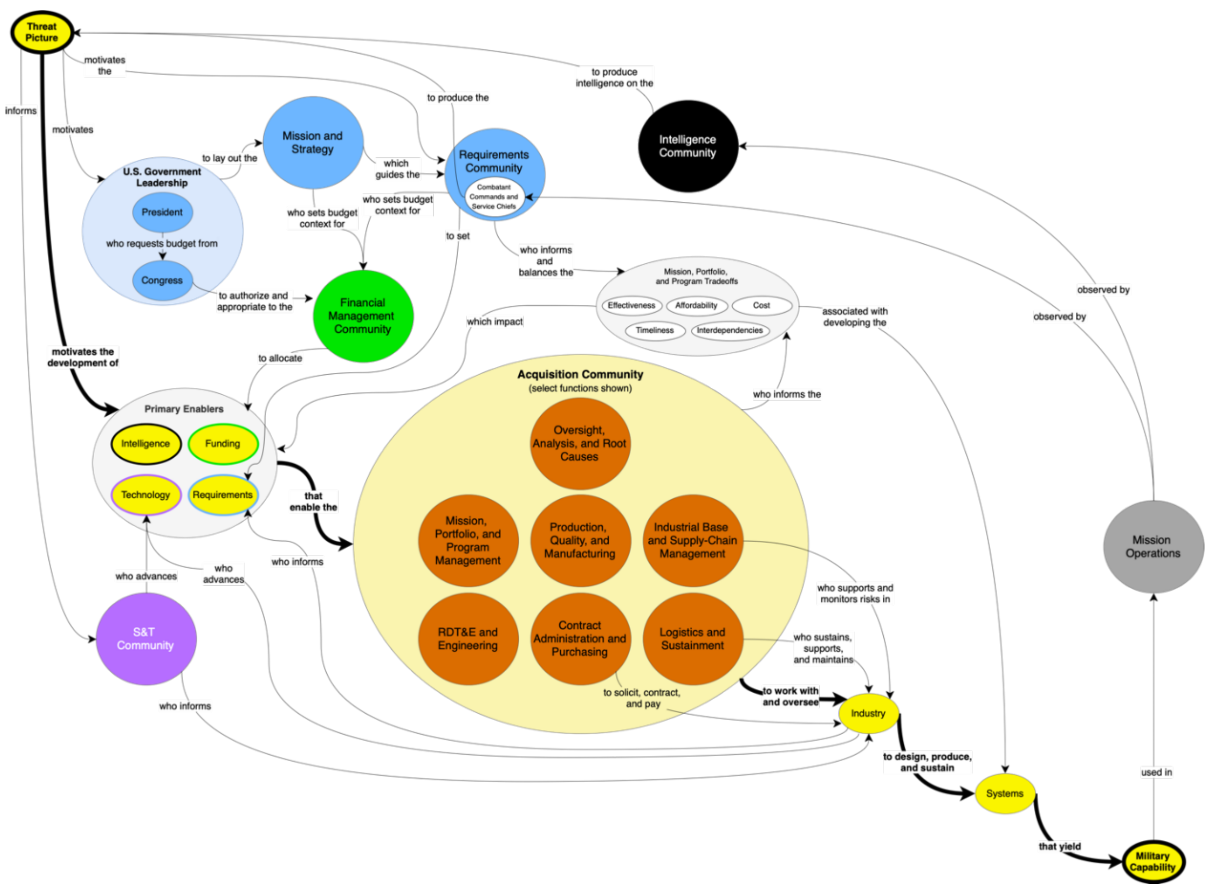


Figure 3. Defense Acquisition Systemigram

Scene 1: Mainstay

The first scene, shown in Figure 4, presents the mainstay of the systemigram, which describes the purpose or main takeaway of the system. The mainstay for this system is that the threat motivates the development of a series of primary enablers (requirements, funding, technology, and intelligence) that enable the Acquisition Community to work with and oversee industry to design, produce, and sustain systems that yield military capability. Military capability, as the last node in the mainstay, is the ultimate goal of the system. Stating this as the ultimate goal was drawn from DoD Directive 5000.01, which states that the acquisition system is designed to deliver "improvements to mission capability" (Under Secretary of Defense for Acquisition and Sustainment [USD(A&S)], 2020a) through the products and services it acquires.

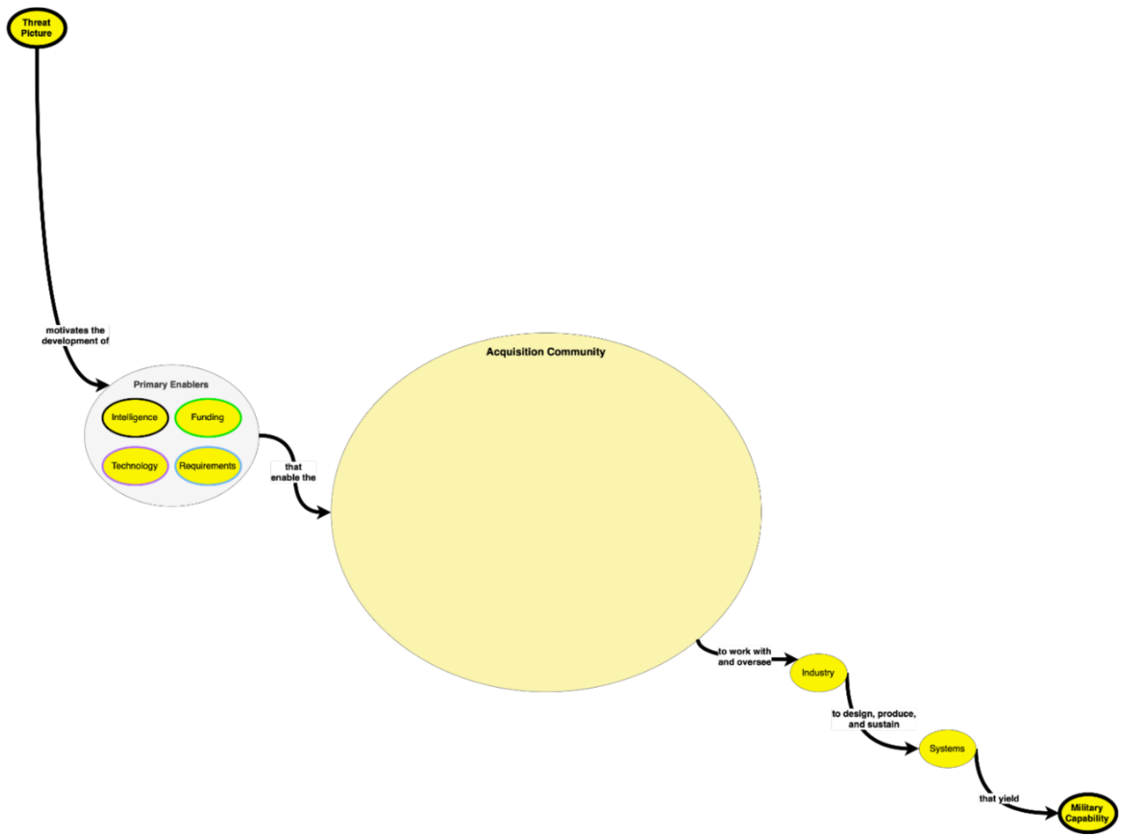


Figure 4. Scene 1: Mainstay

Scene 2: Feedback Loop

The second scene, Figure 5, depicts the feedback loop that exists within the defense acquisition system. The military capability generated by defense acquisition is used in Mission Operations. These operations are then observed by the Intelligence Community and Combatant Commands and Service Chiefs to produce the threat picture.

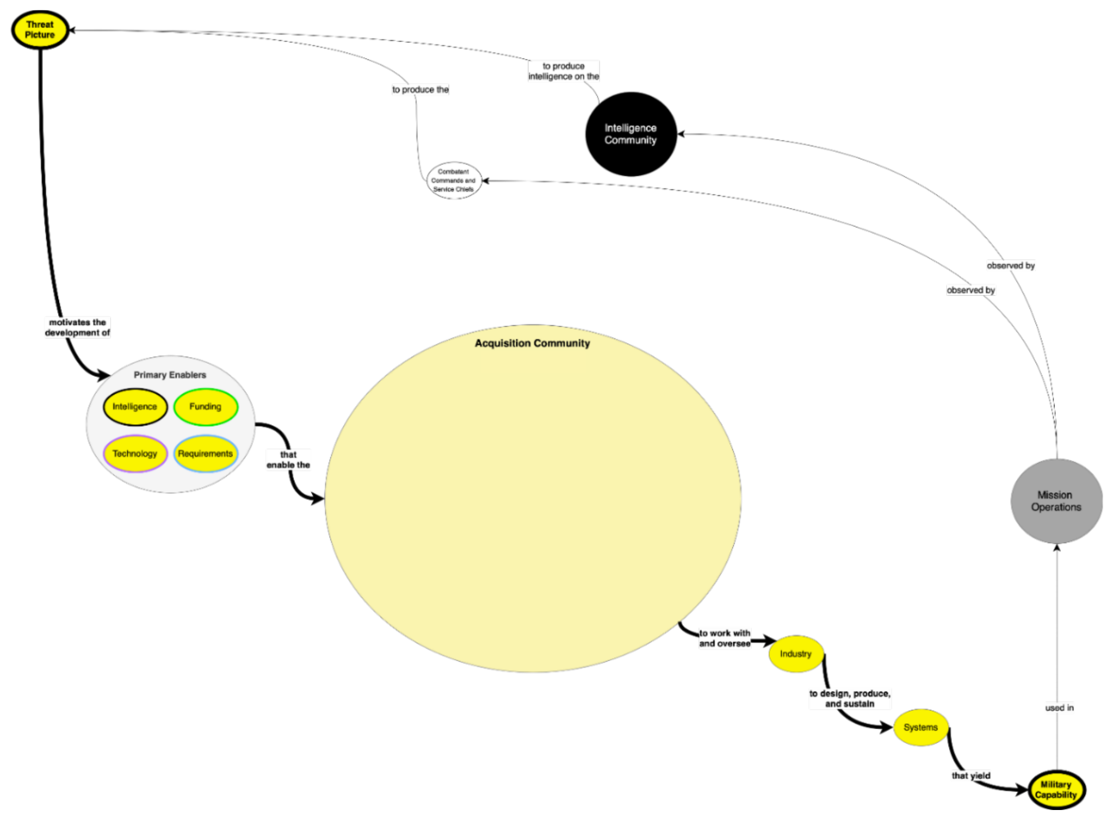


Figure 5. Scene 2: Feedback Loop

Scene 3: Tradeoffs

A key aspect to the overall goal of generating a military capability in a world of limited resources is to produce effective, timely, and affordable solutions (Office of Inspector General, 2020). Performance, schedule, and cost tradeoffs associated with the capability are made by the Service Chiefs and other stakeholders and informed by the Acquisition Community, as seen in Figure 6. These tradeoffs can impact the primary enablers, resulting in changes to requirements and funding adjustments. Note that the Combatant Commands and Service Chiefs, first shown in the second scene, are members of the Requirements Community, but they operate within a larger reporting and budgetary context not explicitly shown in the figure.

The tradeoffs node is labeled "Mission, Portfolio, and Program Tradeoffs" to reflect the increased effort to assess the combined effects of acquired programs as they interact to bring capabilities to the warfighter (Cronk, 2021; GAO, 2007). These assessments and tradeoffs consider how individual acquisition projects fit into larger sets of capabilities (i.e., portfolios) and how they interrelate to serve mission objectives. A mission- and portfolio-based perspective also addresses interdependencies (Cronk, 2021), included in the tradeoffs node of the systemigram, and synchronizes efforts across the entire portfolio, including technologies, capability areas, missions, and programs.

In addition, the tradeoffs node includes reference to both cost and affordability. Cost refers to the price of the system being acquired. Affordability is "the resources projected to be available in the DoD Component portfolio(s) or mission area(s) associated with the program under consideration" (USD[A&S], 2020b). In other words, affordability ensures that the system will be fully funded over its lifetime.

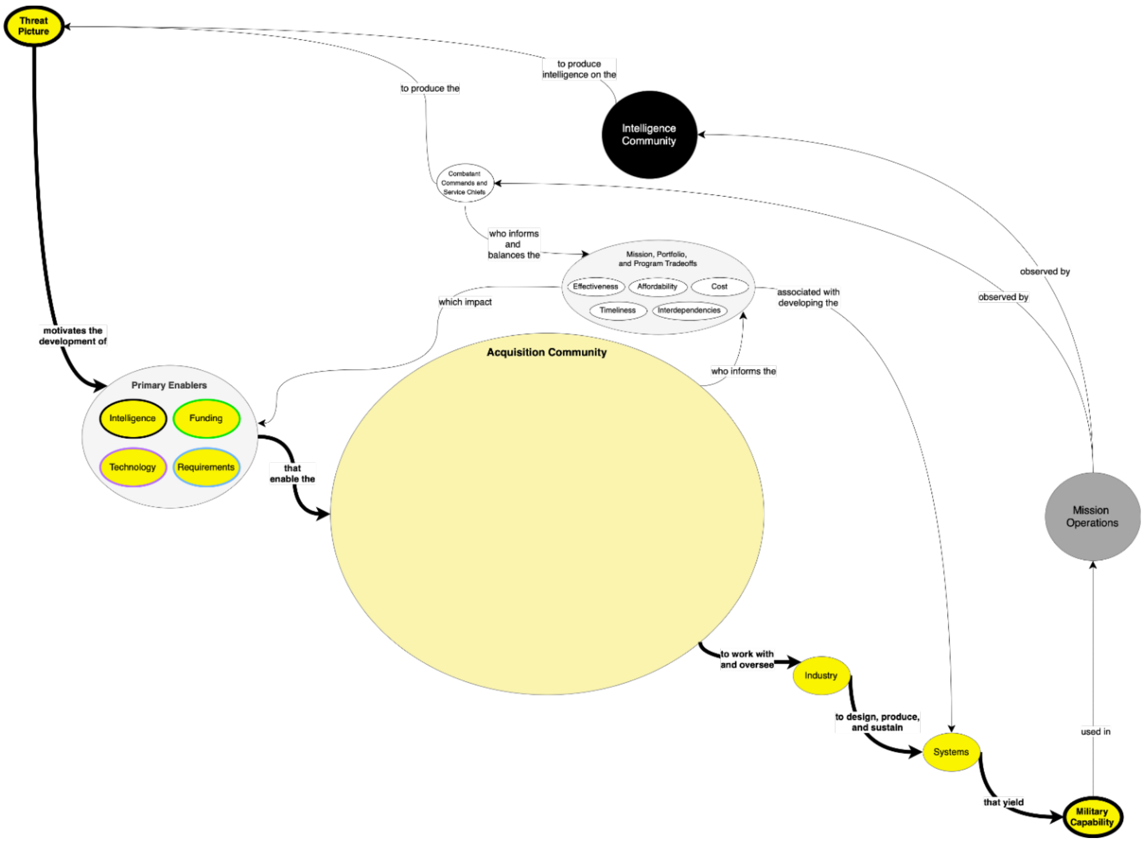


Figure 6. Scene 3: Tradeoffs

Primary Enablers

The primary enablers—requirements, funding, technology, and intelligence—are necessary for acquisition to take place. They can also interact with each other. For example, new technology and usage concepts can drive new requirements, or vice versa. Intelligence drives both requirements and informs acquisition. Funding constrains which needs are deemed important enough to become validated requirements. These interactions are not shown at the level of this diagram, but the presence of the primary enablers indicates their importance in the high-level process.

Scene 4: Requirements

The fourth scene, shown in Figure 7, concerns the primary enabler of *requirements*. The threat picture motivates U.S. Government Leadership (i.e., the President and Congress, in their various roles, responsibilities, and authorities) to lay out the Mission and Strategy, which guides the Requirements Community to set the requirements. Though not explicitly shown in the diagram, the requirements flow through the Acquisition Community and then to industry as system and contractual requirements.

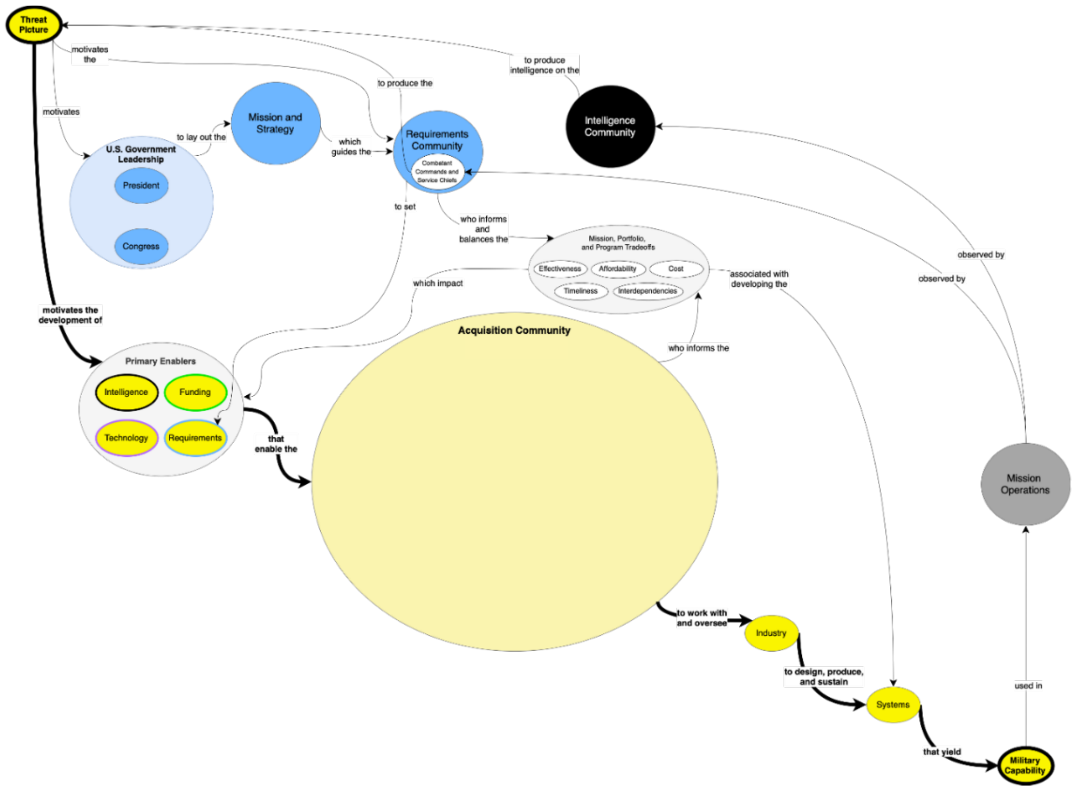


Figure 7. Scene 4: Requirements

Scene 5: Funding

The scene in Figure 8 adds the entities involved in the primary enabler of *funding*. The President requests a budget from Congress, who then authorizes and appropriates a version of the budget back through the Executive Branch hierarchy to the Financial Management Community to allocate funding (initially obligation authority but later as expenditures). The Mission and Strategy and Requirements Communities also set the context for a budget controlled by the Financial Management Community. Given the focus on the acquisition system, the details on how budgets are requested, set, and governed are not shown; that complicated system is out of scope of this paper and deserves its own systemigram (or set of systemigrams).

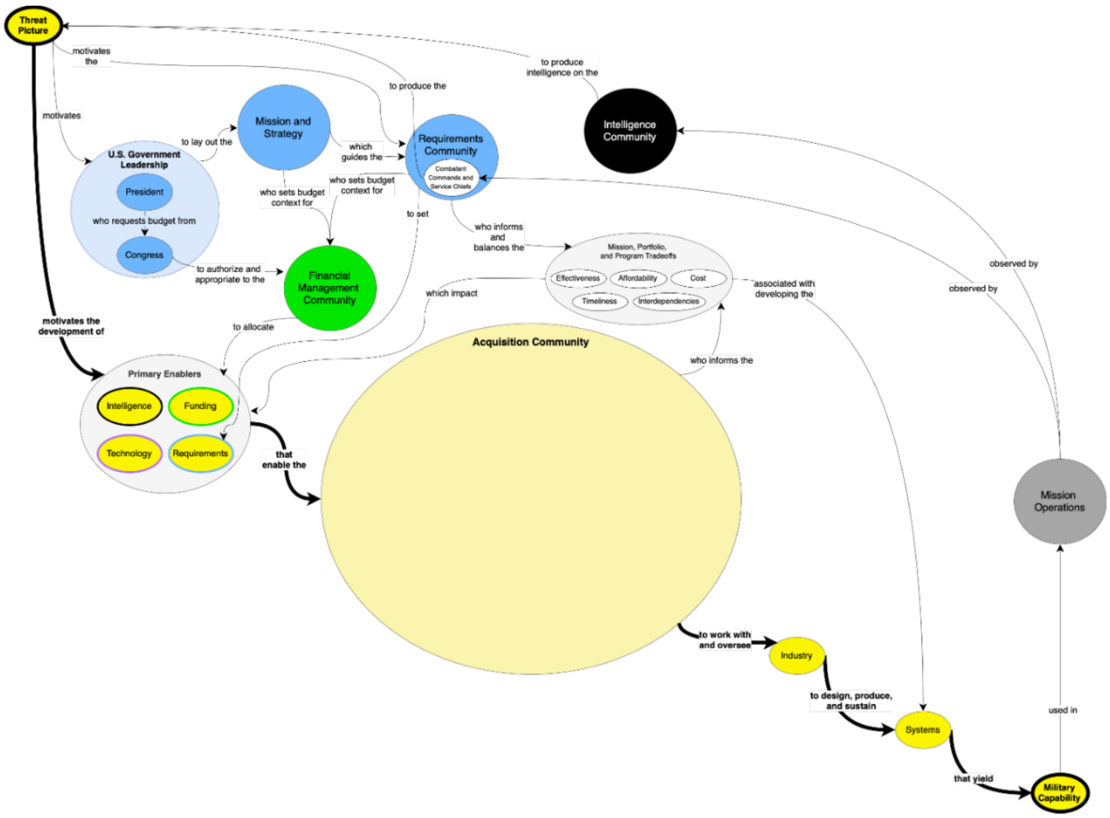


Figure 8. Scene 5: Funding

Scene 6: Technology

Scene 6, illustrated in Figure 9, regards the primary enabler of *technology*. The threat picture informs the Science and Technology (S&T) Community, which then informs industry of specific needs. Here, S&T refers to the applied research efforts within the defense acquisition system. Both S&T and industry advance the state of technology, and research and development of new technologies from industry are a main contributor to the requirements and acquisition processes.

In some cases, the requirements for a project or mission are unable to be fulfilled with technologies that are currently available, resulting in an advancement of the state of technology as a result of the defense acquisition process. In other scenarios, a technological advancement from industry or the research and development efforts within the Acquisition Community triggers the acquisition process. Regardless of whether technological advancement results from or activates the defense acquisition process, as a primary enabler it is required for the acquisition of new systems.

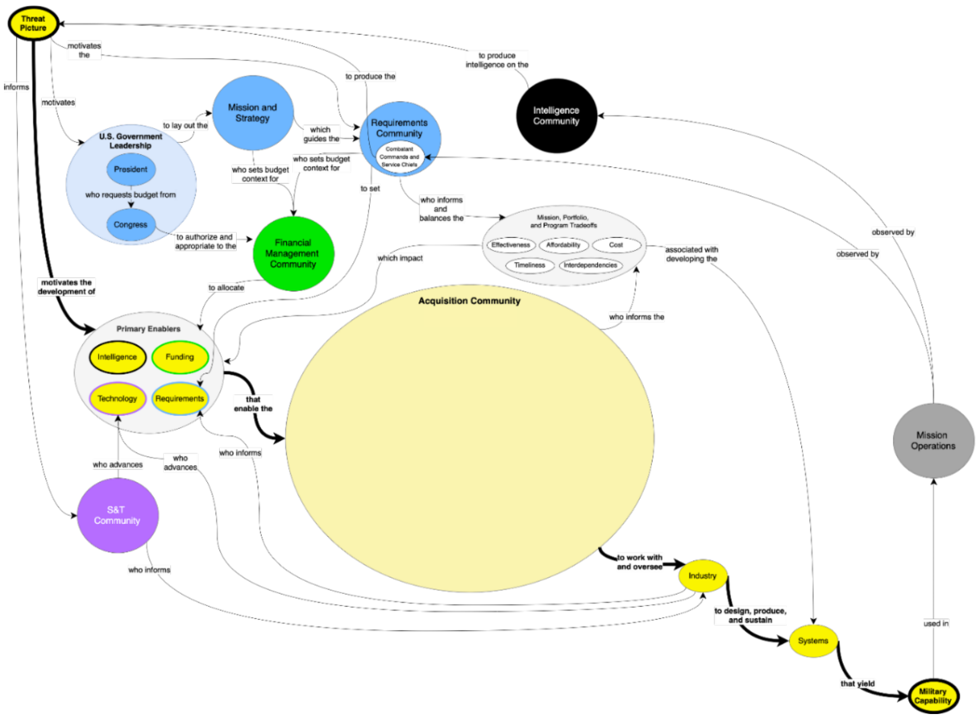


Figure 9. Scene 6: Technology

Scene 7: Acquisition Community

The final scene, Figure 10, fills in some elements of the Acquisition Community. Only select functions have been included as an example of those that exist within the extensive Acquisition Community. The Oversight, Analysis, and Root Causes function serves to examine the successes and deficiencies of the defense acquisition system and their causes. This function monitors the system to ensure that it functions properly and achieves its goal. The function was placed at the top of the Acquisition Community node to emphasize its role in overseeing the workings of the remainder of the community.

Additional functions within the Acquisition Community are Mission, Portfolio, and Program Management; Production, Quality, and Manufacturing; and Research, Development, Test, and Evaluation (RDT&E) and Engineering. Industrial Base and Supply-Chain Management supports and monitors risks in industry, as evidenced by the DoD Office of Industrial Policy's commitment to "providing detailed analyses and in-depth understanding of the increasingly global, commercial, and financially complex industrial supply chain" (Office of Industrial Policy, 2021). Contract Administration and Purchasing solicits, contracts, and pays industry. Logistics and Sustainment sustains, supports, and maintains the systems developed as a result of defense acquisition. Responsibility for sustaining these systems falls on both industry and a member of the Acquisition Community.

Interactions within the Acquisition Community have been intentionally excluded from this systemigram. This better achieves the intention of producing a high-level overview of the acquisition system and how the Acquisition Community interacts with functions and organizations outside of it.

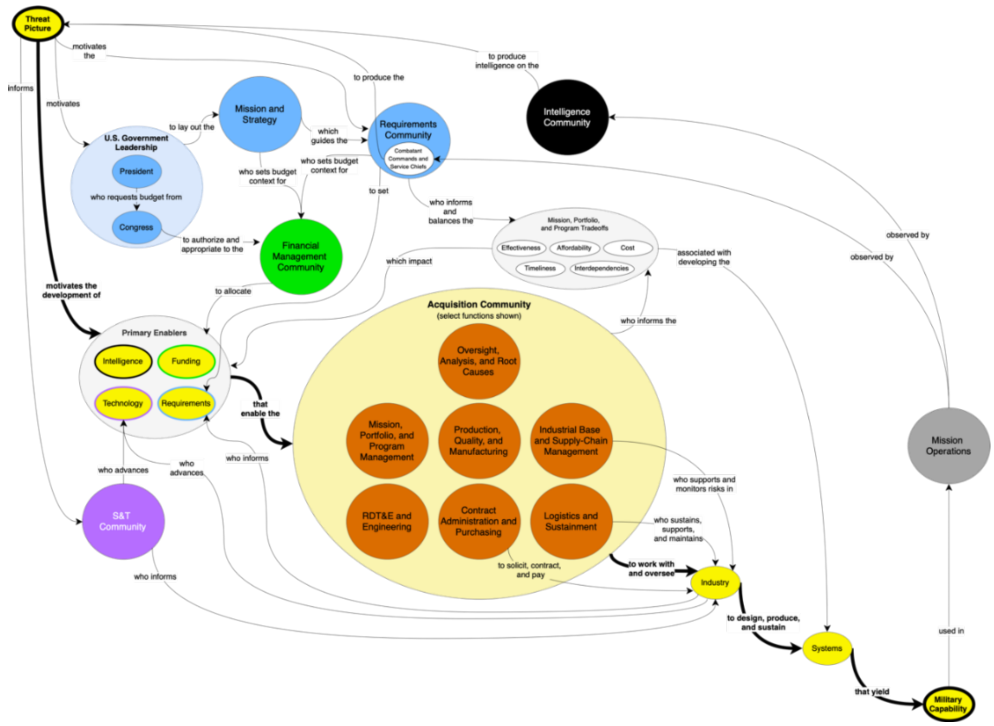


Figure 10. Scene 7: Acquisition Community

Verification and Validation

The verification and validation processes for the Defense Acquisition Systemigram consisted of ensuring that it meets the rules of systemigram creation and that it accurately depicts the defense acquisition system at a high level.

As a Systemigram

While creating the diagram, care was taken to ensure that it meets the rules of systemigram creation. First, the diagram has a mainstay, which reads from the top left to the bottom right (Sauser, 2019). Nodes consist of noun phrases and are not repeated, while arcs comprise verb phrases and do not cross any other arcs in the diagram (Sauser, 2019). Another requirement is that every node has at least one input and output, save for the beginning and ending nodes of the mainstay (Sauser, 2019). Nodes within a containment node (e.g., in this diagram, the primary enablers and nodes within the Acquisition Community) are also exempt from this rule, so long as the containment node has at least one input and output. The Defense Acquisition Systemigram meets this rule, as well. Finally, the imperative of beautification (Sauser, 2019) has been followed through the use of color and line thickness in the diagram. Color aids in understanding the flow of the diagram and visually grouping nodes that appear within the same scene. Line thickness differentiates arcs involved in the mainstay from those that are not. As the rules for creation have been met, the diagram is a legitimate systemigram.

As a High-Level Depiction of Defense Acquisition

To ensure that the diagram accurately depicts the defense acquisition system at a high level, it was shown to another expert within the field. The expert agreed that the diagram would be useful for individuals or universities to better understand the interactions within the defense acquisition system but also expressed concern with keeping it up to date as changes are made to DoD policy. The addition of the Oversight, Analysis, and Root Cause function resulted from this review, to show that the Acquisition Community monitors its process and progress. The goal of the system was also amended to “military capability” from “military advantage” to reflect that the acquisition system still maintains purpose after one threat no longer exists.

Relevance

The relevance of the Defense Acquisition Systemigram includes addressing a gap in the literature, use by individuals within and outside of the defense acquisition system, and the potential to inform systemic improvements for better acquisition outcomes.

Literature Gap

As addressed in the Related Work section, there is a gap in the literature regarding a high-level view of the overall defense acquisition system. This work helps to fill that gap by providing one such view in the form of a systemigram.

The diagram also distinguishes itself from other systemigrams about defense acquisition. The Defense Acquisition Systemigram reflects a functional flow of the system, which differs from the typical process-oriented perspective provided by the systemigram in Figure 1. The Defense Acquisition Systemigram and the diagram in Figure 2 are differentiated by their choices of the beginning node of the mainstay. Figure 2 states that combat veterans drive the defense acquisition system, but the Defense Acquisition Systemigram asserts that it is the threat picture. The systemigram in Figure 2 does not examine the role of intelligence, though it does include the other primary enablers of the Defense Acquisition Systemigram.

Use by Individuals Within and Outside of the System

This systemigram could be useful for both individuals within and outside of the defense acquisition system. While individuals within the system may already possess knowledge of its workings at this high level, the systemigram can stimulate shared understanding and provide useful context about how functions and departments buried within this large enterprise relate to others within the system. Another powerful potential use for the systemigram is to help individuals within the acquisition system refocus on the ultimate goal of achieving military capability, rather than just the immediate objectives of their embedded function and organization.

Similarly, the systemigram can be used by those outside of the defense acquisition system to learn more about how it works. This high-level overview could be useful for professors or students aiming to get involved with defense acquisition, or perhaps Defense Acquisition University students or new members of Congress.

Applications for Acquisition Innovation

This systemigram also has applications for improving defense acquisition. While changes would likely not take place at the level of fidelity shown in the Defense Acquisition Systemigram, it clarifies the key systemic elements and processes involved. For example, the systemigram reinforces that the threat is the reason for the Acquisition Community to exist because it is the beginning node of the mainstay. Also, requirements, funding, intelligence, and technology are necessary enablers of acquisition; impedances in those flows can have negative consequences for the effectiveness of defense acquisition, so simply improving the processes within the Acquisition Community may be insufficient for better outcomes. If not illuminating areas for innovation, at the very least, the diagram could generate discussion about its correctness, which is also useful to gain and codify knowledge about the system.

The Defense Acquisition Systemigram could serve as an “as is” depiction of the system in innovation efforts. Another systemigram, illustrating the “to be” version of the system, could be created and compared to the Defense Acquisition Systemigram. Innovation efforts could then focus on how to achieve the “to be” systemigram from the “as is” systemigram.

A final use of the systemigram is to ensure that changes focus on the ultimate goal of military capability. All decisions and innovation efforts should keep this goal in mind and clarify to stakeholders and employees the importance of their work in relation to this goal.

Future Work

Almost every arc and node in the Defense Acquisition Systemigram could be expanded into its own systemigram to explain the complexity of the relationships in the defense ecosystem. Future projects could work to create a family of systemigrams that more completely illustrate the system and show varying levels of detail. In particular, a systemigram is needed to represent the interactions within the Acquisition Community. These relationships were excluded from the Defense Acquisition Systemigram in order to maintain a high-level overview of the system. The flow of trained personnel throughout the system and the legal advising process were removed from earlier versions of the diagram for a similar reason.

Conclusion

The Defense Acquisition Systemigram represents the interactions between actors in the defense acquisition system at a high level, providing a systems view of the key elements and drivers of the complex system. The diagram addresses a lack of high-level visual representations of the overall system in the literature and also offers applications for individuals within and outside of the system, as well as acquisition innovation. The most profound implication of the Defense Acquisition Systemigram is the ability to recenter the defense acquisition effort on key enablers necessary for acquisition and in creating military capability, which is the ultimate goal of the system.

References

Blair, C. D., Boardman, J. T., & Sauser, B. J. (2007). Communicating strategic intent with systemigrams: Application to the network-enabled challenge. *Systems Engineering*, *10*(4), 309–322.

Cilli, M. V., Parnell, G. S., Cloutier, R., & Zigh, T. S. (2016). A systems engineering perspective on the revised defense acquisition system. *Systems Engineering*, *18*(6), 584–603.

Cronk, T. M. (2021, September 28). *DOD official says concept of integrated deterrence is call to action*. DoD News. https://www.defense.gov/News/News-Stories/Article/Article/2791589/dod-official-says-concept-of-integrated-deterrence-is-call-to-action/

GAO. (2007). *Best practices: An integrated portfolio management approach to weapon system investments could improve DOD's acquisition outcomes* (GAO-07-388). https://www.gao.gov/products/gao-07-388

Mehler, J., McGee, S., & Edson, R. (2010). Leveraging systemigrams for conceptual analysis of complex systems: Application to the U.S. national security system. *8th Conference on Systems Engineering Research*. http://anser.org/docs/asyst-doc/Leveraging-Systemigrams-and-Their-Application-to-the-US-National-Security.pdf

Office of Industrial Policy. (2021). *Home.* DoD. https://www.businessdefense.gov

Office of Inspector General. (2020). *Special report: Lessons learned for Department of Defense acquisition officials during acquisition reform.* DoD. https://media.defense.gov/2020/Aug/04/2002469834/-1/-1/1/DODIG-2020-109.PDF

Sauser, B. (2019, November 18). *Systems thinking workshop* [PowerPoint slides]. Systems Engineering Research Center. https://sercuarc.org/wp-content/uploads/2019/12/Systems-Thinking-Workshop_CombinedSlides.pdf

Under Secretary of Defense for Acquisition and Sustainment. (2020a). *The defense acquisition system* (DoD Directive 5000.01). DoD. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/500001p.pdf?ver=2020-09-09-160307-310

Under Secretary of Defense for Acquisition and Sustainment. (2020b). *Operation of the adaptive acquisition framework* (DoD Instruction 5000.02). DoD. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500002p.pdf

Wade, J., & Batra, K. (2019). *Systems thinking workshop: SystemiTool 2.0* [PowerPoint slides]. Systems Engineering Research Center. https://sercuarc.org/wp-content/uploads/2019/12/Systems-Thinking-Workshop_CombinedSlides.pdf

Determining a Digital Engineering Framework: A Systematic Review of What and How to Digitalize

Stephen M. Waugh— is a recognized expert on the emerging field of digital engineering at the Johns Hopkins University, Applied Physics Laboratory (JHU/APL), and was the acting Chief Scientist of the Combat Systems Group. His current focus is blending the scientific principles of business process management, systems engineering, and decision science, in order to link processes to products to decisions in the Advanced Weapons Systems Group. He served 20 years in the Marine Corps as an AV-8B pilot, rising to the rank of Lieutenant Colonel, and as a Deputy Program Executive Officer at Naval Aviation Systems Command. He holds a BS in Aerospace Engineering from the U.S. Naval Academy, an MBA in Technology Management from University of Phoenix, and is a Doctor of Business Administration candidate at University of Maryland Global Campus. Senator Waugh was a member of the Maryland State Senate from 2014 to 2018, where he passed national landmark legislation on school safety and veterans' issues.

Abstract

This study is a systematic review to determine a conceptual framework for digital engineering, the objective being to select what and how to digitalize Department of Defense (DoD) acquisition processes, data, and decisions. The research question was, What are the best practices for Digitalization and Industry 4.0 to inform DoD acquisition programs? The study analyzed 20 peer-reviewed scholarly articles from the last 5 years, written by academics and practitioners from 19 countries, focused on Digitalization and Industry 4.0 methods and technologies. This study had five major findings: digitalization projects begin with strategic choices; digitalization is done within an ecosystem that constrains the technical options; digitalization requires a method of execution that assesses opportunity and limits risk; digitalization results in new processes using new data models that enable better decisions; feedback on that new business model will come internally from users and externally from customers.

Keywords: digital engineering, digitalization, Industry 4.0, framework, implementation, strategy

Determining A Digital Engineering Framework

The Department of Defense (DoD) published its Digital Engineering (DE) Strategy in 2018. That was followed in 2020 by the Naval Digital Systems Engineering Transformation (DSET) Strategy. Both have the same five goals. The question has arisen of whether or not DE is a new interdisciplinary branch of engineering, like systems engineering is a branch of industrial engineering. At this time, it has no distinct scientific principles applied to build particular things, no unique processes, methods, or protocols; it is only a policy. However, the commercial world embraced Digitalization and Industry 4.0 out of necessity and has realized great opportunities that government can leverage.

Problem Statement

Executing acquisition plans in a predictable, fully resourced manner is challenging (Kraft, 2015). The National Defense Strategy states that greater efficiency in procurement is a national priority (DoD, 2018c). The National Defense Business Operations Plan declares that reforming the business processes is a key strategic goal (DoD, 2018b). The resulting DE Strategy admits that the DoD lags industry on digital transformation solutions (DoD, 2018a).

The DoD DE Strategy has five goals:

1. Formalize the development, integration, and use of models to inform enterprise and program decision-making.
2. Provide an enduring, authoritative source of truth.
3. Incorporate technological innovation to improve the engineering practice.

4. Establish a supporting infrastructure and environments to perform activities, collaborate, and communicate across stakeholders.
5. Transform the culture and workforce to adopt and support DE across the life cycle.

The *Defense Acquisition University Glossary* defines DE as "an integrated digital approach that uses authoritative sources of systems' data and models as a continuum across disciplines to support life cycle activities from concept through disposal" (Defense Acquisition University [DAU], n.d.). However, neither the goals nor the definition answers the critical questions of what or how to implement digitalization.

Rationale

A report by Blackburn et al. (2018) formed the foundation of the DoD DE Strategy, later restated and published in Bone et al. (2019). Neither articulated a conceptual framework for implementation. That is the rationale for this study.

DE discussions often include unfamiliar and somewhat fluid terms. These may include Digital Thread (Kraft, 2020), Digital Twin (Madni et al., 2019), Digital Surrogate (Chakraborty et al., 2021), Electronic Prototype (Rieken et al., 2020), Authoritative Source of Truth (Kraft, 2019), Government Reference Architecture (DoD, 2010), Open Architecture (Keller, 2021), and Agile Software (Scaled Agile, n.d.). This study generally avoids them.

Objective

The objective of this study is to identify the current state of digitalization practices and methods, and to identify a conceptual framework and notional integration of business processes to data products to structured decisions that would satisfy the goals of the DoD DE Strategy. This study is a systematic review.

Potential Significance

Newly digitalized processes would be documented and constrained, with their triggers, inputs, and outputs defined. Policy mandates imposed on a major defense acquisition program (MDAP) would be knowable and trackable over the life cycle of an acquisition program. Program decisions could be made with a common operating picture of the technical and managerial context around a given problem on a variety of levels, in a variety of functions, across the enterprise.

Theoretical Framework

General Systems Theory (von Bertalanffy, 1972) provides a framework that can bridge between systems engineering, business process management, and decision science. A biologist, von Bertalanffy, published his Theory of Organic Shape, "Gestalt," in 1926. He published his view of organisms as physical systems in 1940, and ultimately published the seminal General Systems Theory (von Bertalanffy, 1950). A modern conceptual framework adapted from Marcketti et al. (2009) is shown in Figure 1.

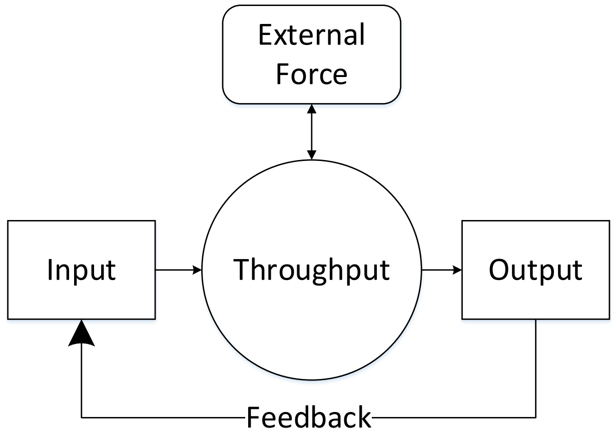


Figure 1. General Systems Theory Conceptual Framework
(Marcketti et al., 2009)

Systems Engineering

The International Council on Systems Engineering (INCOSE) stated that Systems Engineering emerged concurrently with Bertalanffy, at Bell Telephone Labs (INCOSE, n.d.). Hall (1962) defined a methodology for systems engineering to formalize and teach the principles of it. Kossiakoff and Sweet (2003) cited several approaches, including one adopted by the Defense Acquisition University for instruction, shown in Figure 2. It bears note that this engineering process is defined by inputs, a multistep process, outputs, and feedback loops like Systems Theory.

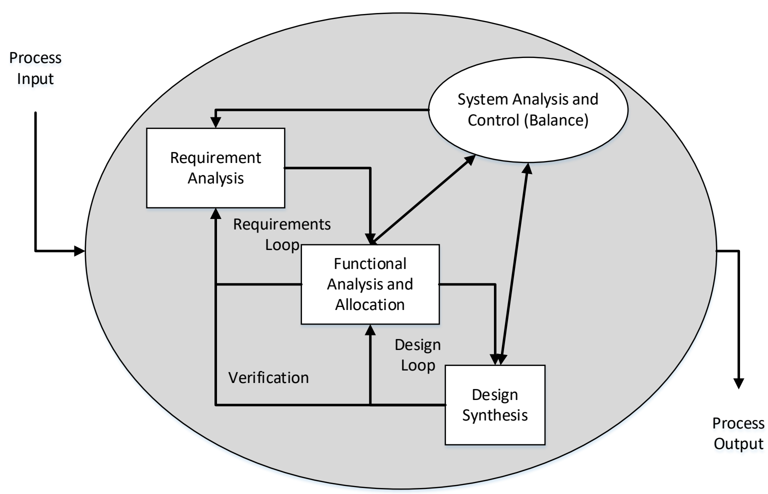


Figure 2. Systems Engineering Process
(DAU, 2001, p. 6)

Business Process Management

Dumas et al. (2013) stated that Business Process Management (BPM) is how work should be performed in order to ensure consistent outputs and to take advantage of improvement opportunities. This includes a circular life cycle of process identification, monitoring, modeling, analysis, and redesign. Business Process Model Notation (BPMN) is the industry standard and is defined by the Object Management Group (OMG), as they do for Systems Modeling Language (SysML). Figure 3 shows the Microsoft Visio default process modeled with inputs, outputs, and feedback loop, also like Systems Theory.

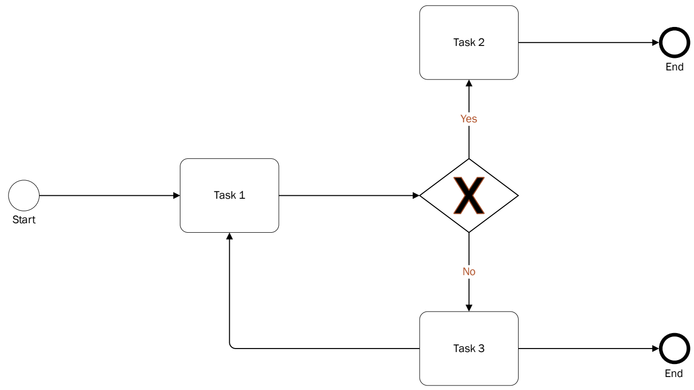


Figure 3. Business Process Model and Notation Example

Note: Default example business process model in Microsoft Visio.

Decision Science

Davis et al. (2005) defined decision science (DS) as human decision-making (why people decide) and the tools that assist it (decision support). Deitrick and Wentz (2015) discussed several theories in DS. They showed that explicit and implicit uncertainty exist throughout the decision process, impacted in part by the changing interactions between steps in a process, the data, and the decision-makers. It bears note that they modeled a decision as a process with input, data, and outputs, as shown in Figure 4, again like Systems Theory.

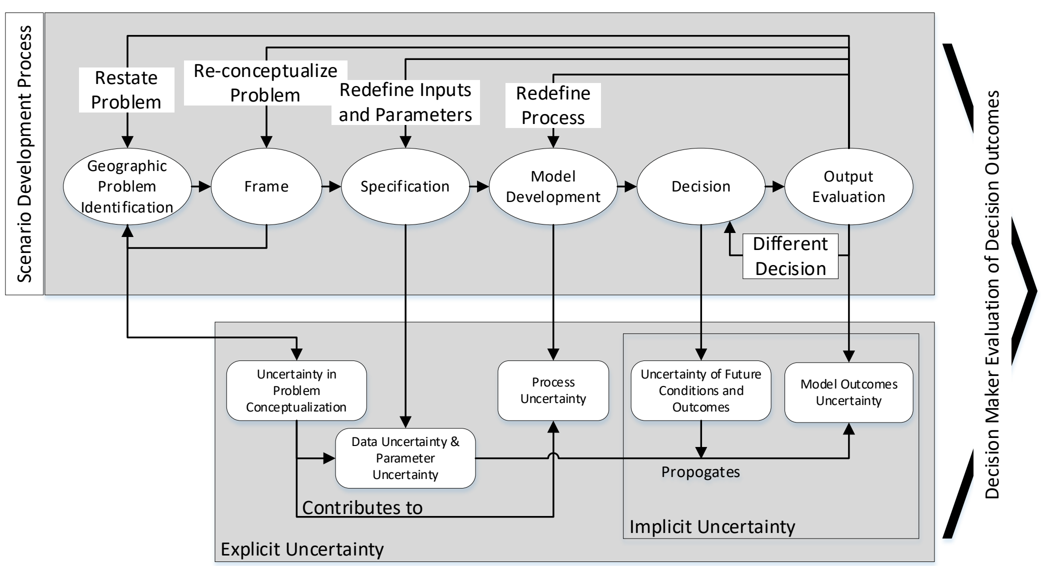


Figure 4. Decision Process Diagram
(Deitrick & Wentz, 2015, p. 548)

Synthesis

General Systems Theory (von Bertalanffy, 1972) is the parent theoretical lens for this study. It describes open systems as organisms that have input, throughput, output, react to external forces, and have a feedback loop. Systems engineering as described by Kossiakoff and Sweet (2003) also had inputs, a process, output, feedback loops, and external forces. Deitrick and Wentz (2015) described decision processes with similar components. Inputs, output, feedback loops, and external forces are the archetypal objects in a BPMN process. Therefore, this framework is a convenient means to bridge these disciplines.

Methodology

General

This study is a systematic review of scholarly journals to provide the evidence-based current state of Digitalization and Industry 4.0 practice and methods. Petticrew and Roberts (2006) described one method of performing a systematic review. Barends et al. (2017) offered a more streamlined approach for narrower questions that require rapid evidence assessments. The generally accepted method for more exhaustive review is the PRISMA, recently updated by Page et al. (2021). The PRISMA Checklist identified 27 items for consideration of inclusion in a systematic review. The research approach for this study was adapted from that and is shown below in Figure 5.

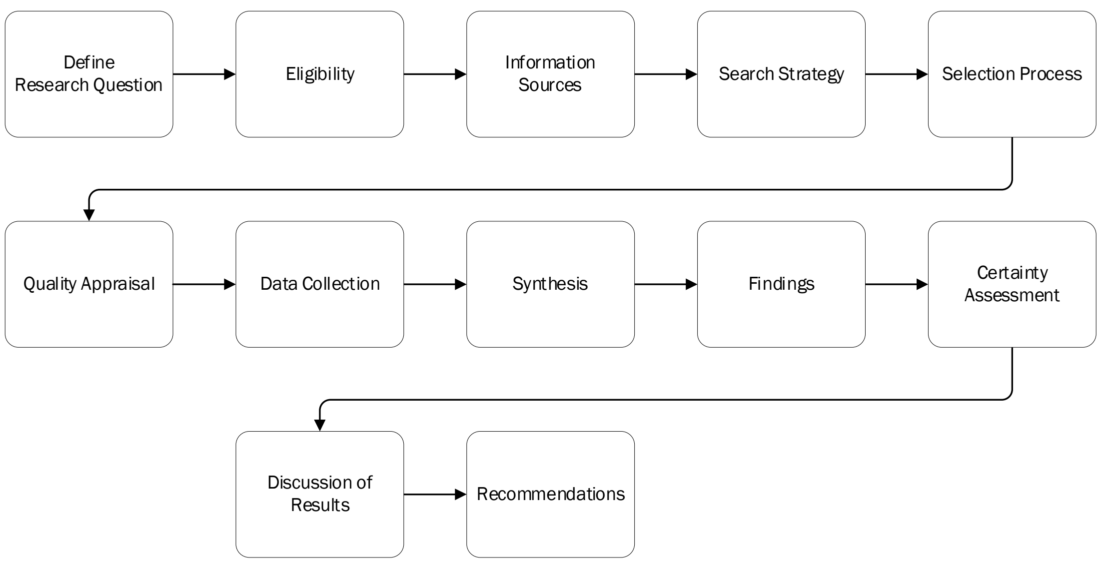


Figure 5. Overview Of Research Approach
(Page et al., 2021)

Research Question

The research question was developed using the Population, Intervention, Comparison, Outcome, Context (PICOC) framework (Barends et al., 2017). Initially, the population was to be the financial technology industry, but scholarly research on that segment proved too narrow, so the population was broadened to general business. The intervention was Digitalization and Industry 4.0 practices and methods, as they are being applied by business operations. Comparison would be to the existing DoD practices. The outcome was improved business processes to create better data products to make better decisions. The context was DoD acquisition programs. Application of PICOC is summarized in Table 1 and led to development of the following research question, *What are the best practices for Digitalization and Industry 4.0 to inform DoD acquisition programs?*

Table 1. PICOC Framework
(Barends et al., 2017)

Population	Who	Commercial Industry
Intervention	What or How	Digitalization and Industry 4.0 practices and methods
Comparison	Basis	Existing DoD processes, data products, and decisions
Outcome	Goal	Better planned and dynamic decision-making, with Lean processes
Context	Circumstance	DoD acquisition program

Eligibility

The inclusion criteria was restricted to peer-reviewed, scholarly journal articles. Only full-text articles were sought. For a fast-moving field, only articles from the last 5 years were accepted. While natural language processing translations provide extraordinary access, English language publications offer less risk of miscommunication. Sources were restricted to journals titled "business" or "management" that had published more than three articles on topic within the last 3 years, demonstrating sustained interest by the publisher, reviewers, authors, and readers.

Information Sources

Google Scholar and ResearchGate were used to conduct initial scoping studies and find preliminary evidence on "digital engineering" that address the current state of digitalization best practices, frameworks, strategies, and implementations, for process, or data, or decisions. The final search of University of Maryland Global Campus (UMGC) OneSearch for evidence was reported using the PRISMA flow diagram (Moher et al., 2009).

Search Strategy

A set of Boolean search terms was developed with the assistance of UMGC librarians. Table 2 explains how they were derived. The final search set was business AND technology, digitization OR digitalization, "best practice" OR framework, strateg* OR implement*, process OR data OR decision.

Table 2. Search Terms and Strings

Concept	Search Term
Technology Industry	business AND technology
Digitalization and Industry 4.0	digitization OR digitalization
Practices and methods	"best practice" OR framework AND strateg* OR implement*
Existing DoD processes, data products, and decisions	process OR data OR decision

Quality Appraisal Tools

For the study to be of value, the source articles must be of quality. Critically appraising data sources prevents information overload, ensures relevance, and is a best practice for evidence-based management (Rousseau, 2006). Weight of Evidence (Gough, 2007) was used to assess the coherence, appropriateness, and relevance of articles. TAPUPAS (Pawson et al., 2003) was used to evaluate the selected articles for transparency, rigor, ethics, and quality assessments for inclusion.

Data Collection

Article meta data extraction involved collecting information such as year of publication, research design, sample size, population (e.g., industry, type of employees), and type of study. Overall trustworthiness was judged. Core data extracted were the explicit findings, discussions, or conclusions of each article.

Synthesis

Collected data were recoded using the theoretical lens of Systems Theory, as relevant to the input, throughput, output, external force, or feedback of the open system. The categorized data were viewed for emergent themes. In the end, inputs clearly shaped strategic decisions, the throughput was the process of digitalization, external forces were part of the ecosystem or technical options, the output was a new business model, and feedback was provided by users and customers.

Findings

Input: Strategy Decisions

Blackburn et al. (2017) studied big data implications on research and development (R&D). They explored three important questions in the degree of change: how would big data refine, innovate, or transform R&D? Those mapped to impacts on strategy, people, technology, and process.

Tortorella et al. (2021) explored the impact of Industry 4.0 on Lean Automation. They found process-oriented technologies had more impact on Lean Production (LP) than product and service technologies. This suggests a choice depending upon the desired target for impact.

Kristoffersen et al. (2020) proposed a Smart Circular Economy for manufacturing companies. This framework translates strategies into business analytics outcomes with digital technologies. It has three major dimensions that are relevant, each with degrees of implementation: Data Transformation, Resource Optimization, and Data Flow Process. This is shown in Figure 6.

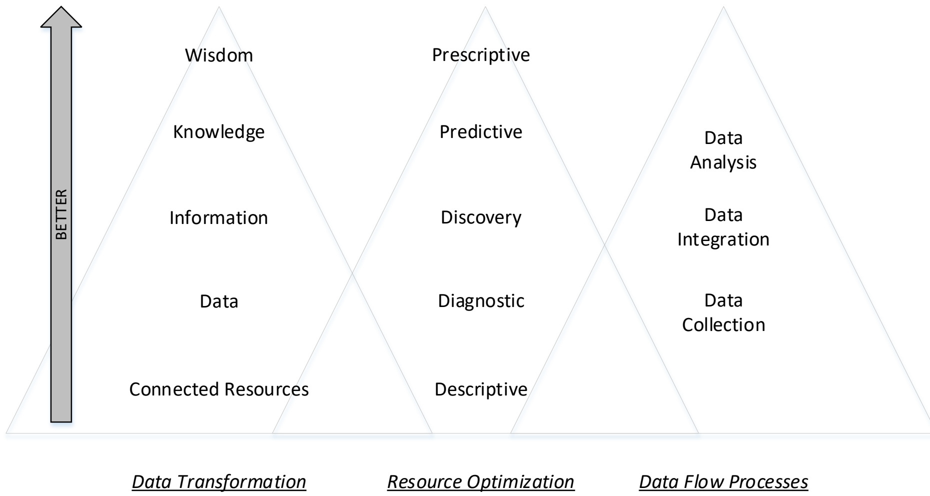


Figure 6. Decision Tool Example
(Kristoffersen et al., 2020, p. 248)

Nosalska et al. (2019) found Industry 4.0 to be a multidimensional system with numerous terms, categories, and variables across its dualistic nature of technical and business. They documented the most common Industry 4.0 design principles over several years of publication. The top recurrent principles were flexibility, real-time capability, decentralization, and modularity.

There were warnings as well. Donnelly (2019) cautioned to avoid over-digitization, while encouraging formal and informal knowledge exchange. This was a key strategic consideration given the tensions of digital transformation.

Throughput: Process

Almost uniformly, the focus was not only on the process of how to digitalize, but emphasizing that business process is the most important target of digitalization. Specifically, the value of digitalization is realized through the transformed underlying business processes (Antonucci et al., 2021). Further, LP is most affected by process technology (Tortorella et al., 2021). Last, process is a critical component of Industry 4.0 implementation in supply chains (Ghadge et al., 2020).

Janiesch et al. (2019) used the 6-step design science research (DSR) process for the design of autonomous agents in the Internet of Things (IoT). They described the DSR steps as (1) Problem Identification, (2) Objectives of a Solution, (3) Design and Development, (4) Demonstration, (5) Evaluation, and (6) Conclusion. They applied this to a scenario of a cyber–physical system (CPS), a self-driving car.

Linde et al. (2021) evaluated opportunities for digital modeling and identified traps to avoid. They found a structured approach for evaluating digital business models had three phases: assessing the opportunity, managing risks, and modeling the future. Concurrently, Linde et al. (2021) found several common traps that must be avoided. First, companies in a rush may not understand the customer value they are creating and fail to satisfy customer needs. Second, not understanding the value delivery process and how the new digitalized process fits within the rest of the corporate context has risks. Last, companies may not understand the new profit formula and means of realizing revenue, simply trusting that digitalization will have made things better.

Output: New Business Model

One critical component of Industry 4.0 implementation is the new digital business model (Ghadge et al., 2020). Another recurring theme is that technical and business-related aspects are interlocked factors (Nosalska et al., 2019). While business model change is enabled by digitalization (Laïfi & Josserand, 2016), the new business model progresses with the business modeling process (Mattsson & Andersson, 2019).

Particularly important for a government procurement agency, Mattsson and Andersson (2019) determined that public–private interaction reveals tensions that drive BPM: structural, behavioral, and organizational. Mattsson and Andersson (2019) concluded a public actor in the complex public network is a much more complex implementation.

External Force: Ecosystem and Technical

There are many external forces to consider. Cong et al. (2021) identified partners as part of the IoT ecosystem. Correani et al. (2020) described a digital transformation ecosystem in which the data platform worked with customers and other players. Dethine et al. (2020) suggested that the ecosystem adapts over time, as did Ghadge et al. (2020). Garay-Rondero et al. (2020) stated that the ecosystem is digital and physical. Gastaldi et al. (2018) considered the firm's larger ecosystem important to a transformation. Linde et al. (2021) described the ecosystem in terms of relationships. Thus the ecosystem could be recoded as people, resources, organization, and supply chain, and a digitalization project will have relationships with all of them.

Ivančić et al. (2019) identified seven main dimensions of digital transformation, including strategy, people, organization, customer, ecosystem, technology, and innovation. In the framework shown in Figure 7, Correani et al. (2020) included data sources, platform, and artificial intelligence (AI). Ghadge et al. (2020) listed data sharing and management as critical.

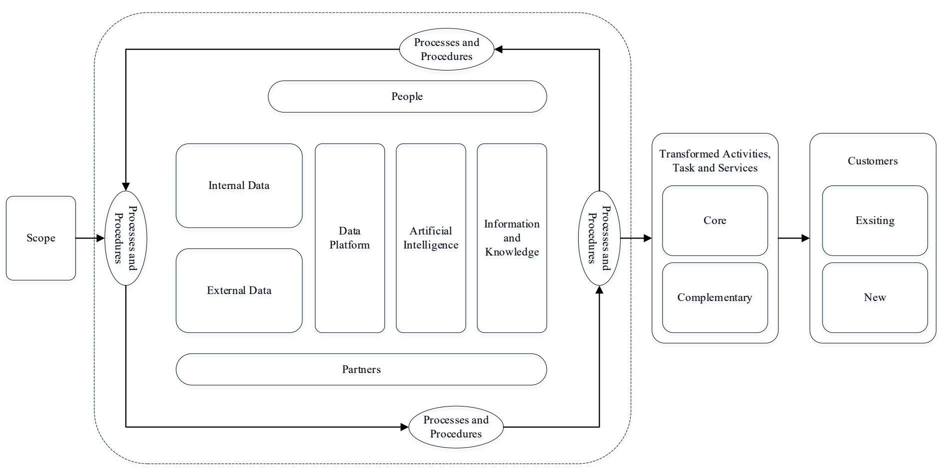


Figure 7. Example Digital Implementation Framework
(Correani et al., 2020, p. 45)

Nosalska et al. (2019) listed many Industry 4.0 key technologies such as CPS, Big Data, IIOT, Cloud Computing/Cloud Manufacturing, Services/Product-as-a-Service/Internet of Services, and System/Architecture. Tortorella et al. (2021) determined that some Industry 4.0 technologies are positively correlated with LP practices, but not all. The emerging technical factors appear to be the platform, technologies, and the data.

Feedback: Users and Customers

Correani et al. (2020) stated that customers could be given immediate feedback if part of the feedback loop. Donnelly (2019) noted opportunity to provide interpersonal feedback to clients and colleagues from digitalization. Garay-Rondero et al. (2020) found feedback to the value chain, and where physical processes affect computations and vice versa. Ghadge et al. (2020) found low customer service could be due to a backlog of feedback on demand. Mattsson and Andersson (2019) found small companies were quicker to adopt platform and content changes based on user feedback, and that customer feedback was important during development.

A Digital Community Infrastructure is digital sharing platforms to share designs and social networks and/or blogs to discuss ideas, questions, and projects (Rieken et al., 2020). IoT technology and Artificial Intelligence of Things (AIoT) empowers the acceleration of digital transformation and real-time collection of data from customers to monitor their conditions or assets to update risk (Cong et al., 2021). Matzler et al. (2018) cautioned that within the existing organization, implementation is highly unlikely to succeed, therefore organizational change is essential to success.

Certainty Assessment

Lewin et al. (2018) described a method of applying the Confidence in Evidence from Reviews of Qualitative Research (CERQual) approach to identify the confidence in findings. CERQual is a framework to evaluate the methods, coherence, adequacy, and relevance of the data used, effectively a self–report card that adds rigor and transparency.

Discussion

Developing the Conceptual Framework

The proposed framework consists of input, throughput, output, feedback, and external forces. In this model, the inputs are the strategic choices to be made for implementation:

1. Degree of Change (Blackburn et al., 2017).
    a. Refine, Innovate, or Transform.
2. Target for Lean Impact (Tortorella et al., 2021).
    a. Process, or Product and Service.
3. Degree of Circular Economy (Kristoffersen et al., 2020).
    a. Data Transformation, Resource Optimization, Data Flow Process.
4. Primary Design Principles (Nosalska et al., 2019).
    a. Flexibility, Real-Time Capability, Decentralization, Modularity.
5. Limit of Digitization (Donnelly, 2019).

These choices should be made with the intent of best achieving the DE Strategy goals of using models to inform decision-making, creating an authoritative source of truth, technological innovation, supporting infrastructure and environments, and transforming the culture and workforce.

Throughput is the process of selecting processes, then digitalizing them. A best practice is to use 6-step design science research process (Janiesch et al., 2019). During execution, evaluate the opportunities and avoid the common traps (Linde et al., 2021). The method of digitally engineering is a process itself. That process is a mini-project plan for each business process under consideration of constraining the problem, setting goals, finding a solution, testing, demonstration, and deployment. Constraining the problem necessarily includes assessing the opportunity for process improvement, because some process improvements may not yield sufficient benefits to make the solutions cost effective, or the margin for improvement may be too small. Last, as the process moves forward the team must continually assess risks. If the project understands the value the process creates (why we do it), the value delivery process (how we do it), and value realization (what we get out of it), those typical traps will be escaped.

The output is a new digitalized business model, where technical and business aspects are intertwined (Nosalska et al., 2019). The more the business model changes, the more the relationships with customers, the supply chain and internal users will change, and new opportunities will arise (Cong et al., 2021; Dethine et al., 2020; Garay-Rondero et al., 2020; Laïfi & Josserand, 2016). This the entire purpose of digitalization. While a DoD acquisition program does not realize revenue (they do not get 'paid' by the Pentagon for systems delivered), they certainly realize costs and deliver product. Having a well-documented business model, especially one that is digitally accessible will enable resource managers to see how their funds are being used, and will enable warfighters to see how their capabilities are being delivered. In addition, legislative authorizers and appropriators will be more easily persuaded to fund programs that are transparent to them.

Many external forces are at work, but they can be grouped into ecosystem constraints and technology opportunities. The ecosystem includes people, resources, organization and the supply chain, which the entity may or may not have control of (Cong et al., 2021; Correani et al., 2020; Dethine et al., 2020; Garay-Rondero et al., 2020; Gastaldi et al., 2018; Linde et al., 2021). Technical forces include the computing environment platforms, technologies, and data (Correani et al., 2020; Ghadge et al., 2020; Ivančić et al., 2019). Technologies do not equally benefit all desired outcomes (Tortorella et al., 2021), but several are key to Industry 4.0 application (Nosalska et al., 2019).

While the number of external forces at work could be infinite, the list must be constrained to provide meaningful decision points. The ecosystem forces were selected because their presence is necessary for success, even if they are constraints beyond the immediate control of the process owner. A process owner may not be able to change the people assigned, or may not have the authority to redirect resources, but both must be present in some limited quantity to succeed. A small operation may have complete control of its organization and culture, while many will be part of a larger organization with a set culture. Both can succeed, but the choices available are different. The digital supply chain for an office is crucial, and every office can identify who it depends on for data to execute owned processes, and what other offices consume data produced. Those players constitute the digital supply chain, and the participation of data suppliers and data consumers in digitally engineering a process is critical. The more they are integrated to the effort, the more opportunities may be exposed for further refinement, enhancing the recursive nature of digitalization.

Technical forces are more likely to be options than constraints. This is where people naturally gravitate to when considering digitalization. An office must consider its computing environment (platform), the technologies available (and affordable), and the data repositories it will require, create, and share. A small office may able to change its platforms, whereas a larger office inside a large organization may have no control, or limited choices within a menu. A major choice will be between on-premises (e.g. desktop) and off-premises (e.g. cloud) computing, and that choice could be driven by security considerations. Industry 4.0 technologies are centered on IoT, and there are many technologies associated with that. Application of technologies like AI, ML, NFC or Bluetooth may accelerate IoT deployment, or they may have limited impact efficacy; being judicious is important. The new business model will hinge on the new data model. Businesses can collect data they never use, or fail to relate or visualize the data they have in a usable manner. A vast repository of stove piped data serves nobody. Data that is interrelated cross-functionally is more likely to have meaning. Data should be collected, created and shared because it is required to execute a process or make a decision.

Feedback will come first from internal users, eventually from external customers, as well as the digital supply chain (Cong et al., 2021; Garay-Rondero et al., 2020; Rieken et al., 2020). Communication with them is essential to success and subsequent adjustments. Providing a means for users to provide faster feedback via a Digital Community Infrastructure will lead to changes in the organizational culture and increase likelihood of acceptance, as users feel they are an integral part of changing the way they do their work. If feedback from those users is not aggressively sought, there is a risk they will obstruct change or sabotage the project. Those users must include not only the performers of a given process, but the users of its results, the decision-makers. The best process with poor visualizations may not improve outcomes.

Figure 8 illustrates the derived conceptual framework for DE. DoD goals feed project strategy decisions, the ecosystem constrains technology choices, process defines execution, a new business model delivers efficiencies, and feedback informs recursion.

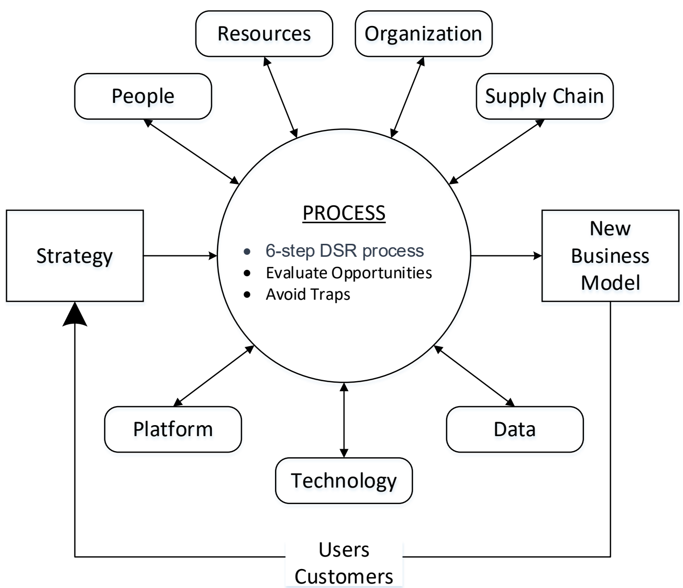


Figure 8. Digital Engineering Conceptual Framework

The goal of digitalization is to arrive at a new set of processes that use a new set of data to achieve value. It is easy to see digitalization merely as a problem of new applications, or the introduction of AI into processes, or new data models depending upon personal perspective or experience. However, none of those solutions alone will have sustained or meaningful impact. New models may be better but may not result in better decisions if disconnected from a unified data model. A web services firm may be able to house petabytes of data for decades, but if it is not designed for people to use with their digital supply chain, its customer value is limited. Using AI as support infrastructure to communicate with customers is common, but without integration with the business process it may not deliver value.

Entities have known they should digitalize but did not know what or how to implement it. This framework provides a means to choose what projects to do and how to execute them in a balanced way.

Recommendations

Establish the implementation framework. Decide what external forces are strengths, weaknesses, opportunities, or threats. To achieve DE goals, decide the strategy. Determine the desired degree of change, impact target, circular economy, design principles, and delimit the changeable processes. Model those processes as-is, to-be, and assess risk, as part of a disciplined project plan. Engineer a new data model on the proper platform with select technology, fed by new processes, and feeding others internally and externally. Communicate with the affected users, customers, and suppliers continuously, seeking failure early and rewarding good outcomes. Plan on necessary organizational changes. Monitor changes to the business model; prepare to adjust.

Limitations, Implications, and Risk

This systematic review was streamlined for rapid completion. While the search was conducted on UMGC library databases, a significant number of results were excluded based solely on the title or abstract, and may be subject to selection bias (Nunan et al., 2017). The search terms may be subject to selection bias. Article content may have been ignored or highlighted based on the author's experience, injecting confirmation bias (Spencer et al., 2018). Digitalization is a rapidly evolving practice with hotly competing providers who need a proprietary edge, which resists scholarly publication.

Engineering is commonly defined as the application of scientific principles to build things. The branches and subbranches are differentiated by using particular scientific principles to build particular things: this is what differentiates mechanical engineering from software engineering. This paper associates the principles of systems engineering, business process management, and decision science for the purposes of describing a DE framework.

The Accreditation Board for Engineering and Technology (ABET) certifies more than 3,000 programs at over 600 U.S. institutions with 75 engineering programs, yet none are "digital engineering." DE is not currently a defined branch of engineering; therefore, few journal articles reference it. DE might be a subbranch sibling of Systems Engineering if a distinct DE process is proposed and accepted.

Entities have known they should digitalize but did not know what or how to implement it. This framework provides a means to choose those projects and execute them in a balanced way. This framework is being deployed in a case study integrated product team (IPT) this summer.

According to Matzler et al. (2018), the biggest risk is the existing organization; therefore, companies need a new culture, with great incentives to innovate and small penalties for mistakes. Failing faster, cheaper, will lead to success in digital transformation.

Conclusion

Answer to the Research Question

The research question was: What are the best practices for Digitalization and Industry 4.0 to inform DoD acquisition programs? The study found broadly that an implementation framework is necessary to properly apply Industry 4.0 technology to the digitalization of business processes. In the case of the DoD, the proposed framework shows DE Strategy goals guide implementation decisions, the ecosystem constrains technology choices, an executable process is defined, the resulting new business model delivers efficiencies, and feedback informs recursion.

A conceptual framework was proposed that integrates these elements, as an evidence-based recommendation. A DoD agency that applied this method would be a cutting edge digitally engineered entity, capable of continuous digital evolution.

References

References marked with an asterisk (*) indicate studies included in the systematic review.

*Antonucci, Y. L., Fortune, A., & Kirchmer, M. (2021). An examination of associations between business process management capabilities and the benefits of digitalization: All capabilities are not equal. *Business Process Management Journal*, *27*(1), 124–144. https://doi-org.ezproxy.umgc.edu/10.1108/BPMJ-02-2020-0079

Barends, E., Rousseau, D. M., & Briner, R. B. (Eds.). (2017). *CEBMa guidelines for rapid evidence assessments in management and organizations, version 1.0*. Center for Evidence-Based Management. https://www.cebma.org/wp-content/uploads/CEBMa-REA-Guideline.pdf

*Blackburn, M., Alexander, J., Legan, J. D., & Klabjan, D. (2017). Big data and the future of R&D management: The rise of big data and big data analytics will have significant implications for R&D and innovation management in the next decade. *Research Technology Management*, *60*(5), 43–50. https://doi-org.ezproxy.umgc.edu/10.1080/08956308.2017.1348135

Blackburn, M., Verma, D., Dillon-Merrill, R., Blake, R., Bone, M., Chell, B., Dove, R., Dzielski, J., Grogan, P., Hoffenson, S., Hole, E., Jones, R., Kruse, B., Pochiraju, K., Snyder, C., & Cloutier, R. (2018). *Transforming systems engineering through model-centric engineering* (A013 Interim Technical Report SERC-2017-TR-111). Stevens Institute of Technology, Systems Engineering Research Center. https://apps.dtic.mil/sti/citations/AD1058354

Bone, M. A., Blackburn, M. R., Rhodes, D. H., Cohen, D. N., & Guerrero, J. A. (2019). Transforming systems engineering through digital engineering. *Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, *16*(4), 339–355. https://www.doi.org/10.1177/1548512917751873

Chakraborty, S., Adhikari, S., & Ganguli, R. (2021). The role of surrogate models in the development of digital twins of dynamic systems. *Applied Mathematical Modelling*, *90*, 662–681. https://doi-org.ezproxy.umgc.edu/10.1016/j.apm.2020.09.037

*Cho, Y., & Lee, Y. (2016). The use of IP profiles in selecting and structuring R&D alliances. *Research Technology Management*, *59*(2), 18–27. https://doi-org.ezproxy.umgc.edu/10.1080/08956308.2015.1137191

*Cong, L. W., Li, B., & Zhang, Q. T. (2021). Internet of things: Business economics and applications. *Review of Business*, *41*(1), 15–29.

*Correani, A., DeMassis, A., Frattini, F., Petruzzelli, A. M., & Natalicchio, A. (2020). Implementing a digital strategy: Learning from the experience of three digital transformation projects. *California Management Review*, *62*(4), 37–56. https://doi.org/10.1177/0008125620934864

Davis, P. K., Kulick, J., & Egner, M. (2005). *Implications of modern decision science for military decision-support systems* (Report No. MG-360-AF). RAND. https://www.rand.org/pubs/monographs/MG360.html

Defense Acquisition University. (n.d.). Digital engineering. In *Defense Acquisition University glossary*. Retrieved May 20, 2021, from https://www.dau.edu/glossary/Pages/GlossaryContent.aspx?itemid=27345#:~:text=An%20integrated%20digital%20approach%20that,activities%20from%20concept%20through%20disposal

Defense Acquisition University. (2001). *Systems engineering fundamentals.* https://acqnotes.com/wp-content/uploads/2017/07/DAU-Systems-Engineering-Fundamentals.pdf

Deitrick, S., & Wentz, E. A. (2015). Developing implicit uncertainty visualization methods motivated by theories in decision science. *Annals of the Association of American Geographers*, *105*(3), 531–551. https://www.doi.org.ezproxy.umgc.edu/10.1080/00045608.2015.1012635

DoD (2010). *Reference architecture description*. https://www.dragon1.com/downloads/DoD-ref-architecture.pdf

DoD. (2018a). *Digital engineering strategy*. https://ac.cto.mil/digital_engineering

DoD. (2018b). *National defense business operations plan*. https://cmo.defense.gov/Publications/NDBOP.aspx

DoD. (2018c). *National defense strategy*. https://www.defense.gov/Explore/Spotlight/National-Defense-Strategy

*Dethine, B., Enjolras, M., & Monticolo, D. (2020). Digitalization and SMEs' export management: Impacts on resources and capabilities. *Technology Innovation Management Review*, *10*(4), 18–34. https://doi-org.ezproxy.umgc.edu/10.22215/timreview/1344

*Donnelly, R. (2019). Aligning knowledge sharing interventions with the promotion of firm success: The need for SHRM to balance tensions and challenges. *Journal of Business Research*, *94*, 344–352. https://doi-org.ezproxy.umgc.edu/10.1016/j.jbusres.2018.02.007

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. (2013). *Fundamentals of business process management*. Springer-Verlag Berlin Heidelberg. https://www.doi.org/10.1007/978-3-642-33143-5_1

*Garay-Rondero, C. L., Martinez-Flores, J. L., Smith, N. R., Caballero Morales, S. O., & Aldrette-Malacara, A. (2020). Digital supply chain model in Industry 4.0. *Journal of Manufacturing Technology Management*, *31*(5), 887–933. https://doi-org.ezproxy.umgc.edu/10.1108/JMTM-08-2018-0280

*Gastaldi, L., Appio, F. P., Corso, M., & Pistorio, A. (2018). Managing the exploration-exploitation paradox in healthcare: Three complementary paths to leverage on the digital transformation. *Business Process Management Journal*, *24*(5), 1200–1234. https://doi-org.ezproxy.umgc.edu/10.1108/BPMJ-04-2017-0092

*Ghadge, A., Er Kara, M., Moradlou, H., & Goswami, M. (2020). The impact of Industry 4.0 implementation on supply chains. *Journal of Manufacturing Technology Management*, *31*(4), 669–686. https://doi-org.ezproxy.umgc.edu/10.1108/JMTM-10-2019-0368

Gough, D. (2007). Weight of evidence: A framework for the appraisal of the quality and relevance of evidence. *Applied and Practice-Based Research*, *22*(2), 213–228. https://www.doi.org/10.1080/02671520701296189

Hall, A. (1962). *A methodology for systems engineering*. D. Van Nostrand Company.

INCOSE. (n.d.). *History of systems engineering*. Retrieved March 30, 2022, from https://www.incose.org/about-systems-engineering/history-of-systems-engineering

*Ivančić, L., Vukšić, V.B. & Spremić, M. (2019). Mastering the digital transformation process: Business practices and lessons learned. *Technology Innovation Management Review*, *9*(2), 36–50. https://doi-org.ezproxy.umgc.edu/10.22215/timreview/1217

*Janiesch, C., Fischer, M., Winkelmann, A., & Nentwich, V. (2019). Specifying autonomy in the Internet of Things: The autonomy model and notation. *Information Systems & E-Business Management*, *17*(1), 159–194. https://doi-org.ezproxy.umgc.edu/10.1007/s10257-018-0379-x

Keller, J. (2021). The latest trends in power electronics: It's not just about stand-alone devices anymore, as power systems designers tackle power density, thermal management, open-architecture standards, systems integration, and obsolescence management. *Military & Aerospace Electronics*, *32*(4), 23–29.

Kossiakoff, A., & Sweet, W. (2003). *Systems engineering principles and practice.* Wiley and Sons.

Kraft, E. M. (2015, January 5–9). *HPCMP CREATE-AV and the Air Force digital thread* [Paper presentation]. AIAA SciTech Forum, Kissimmee, FL.

Kraft, E. M. (2019, January 7–9). *Value-creating decision analytics in a lifecycle digital engineering environment* [Paper presentation]. AIAA SciTech Forum, San Diego, CA. https://doi.org/10.2514/6.2019-1364

Kraft, E. M. (2020, January 5). *Digital engineering enabled systems engineering performance measures* [Paper presentation]. AIAA Scitech 2020 Forum, Orlando, FL. https://www.doi.org/10.2514/6.2020-0552

*Kristoffersen, E., Blomsma, F., Mikalef, P., & Li, J. (2020). The smart circular economy: A digital-enabled circular strategies framework for manufacturing companies. *Journal of Business Research*, *120*, 241–261. https://doi-org.ezproxy.umgc.edu/10.1016/j.jbusres.2020.07.044

*Laïfi, A., & Josserand, E. (2016). Legitimation in practice: A new digital publishing business model. *Journal of Business Research*, *69*(7), 2343–2352. https://doi-org.ezproxy.umgc.edu/10.1016/j.jbusres.2015.10.003

Lewin, S., Booth, A., Glenton, C., Munthe-Kaas, H., Rashidian, A., Wainwright, M., Bohren, M. A., Tunçalp, Ö, Colvin, C. J., Garside, R., Carlsen, B., Langlois, E. V., & Noyes, J. (2018). Applying GRADE-CERQual to qualitative evidence synthesis findings: Introduction to the series. *Implementation Science*, *13*(Suppl. 1). https://doi.org/10.1186/s13012-017-0688-3

*Linde, L., Sjödin, D., Parida, V., & Gebauer, H. (2021). Evaluation of digital business model opportunities: A framework for avoiding digitalization traps. *Research Technology Management*, *64*(1), 43–53. https://doi-org.ezproxy.umgc.edu/10.1080/08956308.2021.1842664

Madni, A. M., Madni, C. C., & Lucero, S. D. (2019). Leveraging digital twin technology in model-based systems engineering. *Systems*, *7*(1), 7.

Marcketti,S., Niehm,L., & Fuloria, R. (2009). An exploratory study of lifestyle entrepreneurship and its relationship to life quality. *Family and Consumer Sciences Research Journal*, *34*, 241–259. https://www.doi.org/10.1177/1077727X05283632
*Mattsson, L.-G., & Andersson, P. (2019). Private-public interaction in public service innovation processes: Business model challenges for a start-up EdTech firm. *Journal of Business & Industrial Marketing*, *34*(5), 1106–1118. https://doi-org.ezproxy.umgc.edu/10.1108/JBIM-10-2018-0297
*Matzler, K., Friedrich von den Eichen, S., Anschober, M., & Kohler, T. (2018). The crusade of digital disruption. *Journal of Business Strategy*, *39*(6), 13–20. https://doi-org.ezproxy.umgc.edu/10.1108/JBS-12-2017-0187
Moher, D., Liberati, A., Tetzlaff, J., Altman D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097
*Nosalska, K., Piątek, Z. M., Mazurek, G., & Rządca, R. (2019). Industry 4.0: Coherent definition framework with technological and organizational interdependencies. *Journal of Manufacturing Technology Management*, *31*(5), 837–862. https://doi-org.ezproxy.umgc.edu/10.1108/JMTM-08-2018-0238
Nunan D., Bankhead C., & Aronson, J. K. (2017). *Selection bias*. Catalogue Of Bias. http://www.catalogofbias.org/biases/selection-bias/
Object Management Group. (2014). *OMG business process model and notation (OMG BPMN™) version 2.0.2*. https://www.omg.org/spec/BPMN
Object Management Group. (2019). *OMG systems modeling language (OMG SysML™) version 1.6*. https://www.omg.org/spec/SysML/1.6
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., & Mulrow, C. D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal*, *372*(71), 1–9. https://www.doi.org/10.1136/bmj.n71
Pawson, R., Boaz, A., Grayson, L., Long, A., & Barnes, C. (2003). *Types and quality of social care knowledge. Stage two: Towards the quality assessment of social care knowledge.* Social Care Institute for Excellence.
Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Blackwell Publishing Company.
*Rieken, F., Boehm, T., Heinzen, M., & Meboldt, M. (2020). Corporate makerspaces as innovation driver in companies: A literature review-based framework. *Journal of Manufacturing Technology Management*, *31*(1), 91–123. https://doi-org.ezproxy.umgc.edu/10.1108/JMTM-03-2019-0098
Rousseau, D. M. (2006). Is there such as thing as "evidence based management"? *Academy of Management Review*, *31*(2), 256–269. https://doi.org/10.5465/amr.2006.20208679
Scaled Agile. (n.d.). *SAFe.* Retrieved March 31, 2022, from https://www.scaledagileframework.com
Spencer, E. A., & Heneghan, C. (2018). *Confirmation bias*. Catalogue of Bias. www.catalogueofbiases.org/biases/confirmationbias
*Tortorella, G., Sawhney, R., Jurburg, D., de Paula, I. C., Tlapa, D., & Thurer, M. (2021). Towards the proposition of a Lean Automation framework: Integrating Industry 4.0 into Lean Production. *Journal of Manufacturing Technology Management*, *32*(3), 593–620. https://doi-org.ezproxy.umgc.edu/10.1108/JMTM-01-2019-0032
von Bertalanffy, L. (1926). Untitled. *Wilhelm Roux' Archiv Fur Entwicklungsmechanik Der Organismen*, *108*(2), 413–416. https://doi-org.ezproxy.umgc.edu/10.1007/BF02080842
von Bertalanffy, L. (1940). Der organismus als physikalisches system betrachtet. *The Science of Nature*, *28*(33), 521–531.
von Bertalanffy, L. (1950). An outline of general system theory. *British Journal for the Philosophy of Science*, *1*, 134–165.
von Bertalanffy, L. (1972). The history and status of general systems theory. *The Academy of Management Journal*, *15*(4), 407–426. https://www.doi.org/10.2307/255139

Instrumenting the Acquisition Design Process: Developing Methods for Engineering Process Metrics Capture and Analysis

Lorri Bennett—is a research engineer and the Head of the Product and Process Design Department in the Manufacturing Systems Division at the Applied Research Lab (ARL) at Penn State. Her specialty is in development of software tools and methods which support Set Based Design and Model Based System Engineering. For the past 20+ years, she has focused on the development of conceptual design environments for complex engineering systems, performing trade studies in support of these applications, and advanced data exploration methods. She has experience in a broad range of application areas, including undersea vehicles, undersea weapons, manned submersibles, unmanned aerial systems, and ground combat vehicle systems. She has been the project lead on numerous government-sponsored research projects, including efforts for NAVSEA, the Office of Naval Research, and the Marine Corps. For the past several years, she has performed as the Lead Systems Engineer and a research engineer on the advanced lightweight torpedo program that is being executed by ARL and funded by Program Executive Office Submarines. Bennett has a Bachelor of Science degree in mechanical engineering from Penn State University. [lab27@arl.psu.edu]

Dr. Jay Martin—is a Research Associate at the Applied Research Laboratory. He joined the Manufacturing Technology Division at the Applied Research Laboratory in 1990. Dr. Martin was the principal investigator for the ONR undersea weapons MSDO research program for 5 years. He was the principal investigator for Lightweight Torpedo Technology (LTT) Insertion FNC program. Dr. Martin was a co-principal investigator in 3-year effort to create a set of tools to perform trade space exploration for electric torpedoes and a 2-year effort to develop trade study tool to explore the hydrodynamic properties of torpedo and UUV-like systems. Dr. Martin was the principal investor of an ONR basic research project to define research topics in Total Ownership Cost. He also supported a Marine Corps life-cycle cost estimating effort called Framework for Assessing Cost and Technology (FACT). Currently, Dr. Martin is the lead researcher in current efforts in Model-Based Engineering on underwater vehicles at ARL. Dr. Martin has over 10 referred conference proceeding and eight journal articles principally on the topic of using kriging models (Gaussian Process Models) for response surface interpolation of data sets, dimension reduction, and uncertainty quantification. [jdm111@arl.psu.edu]

Dr. Simon Miller—studied mechanical engineering at The Pennsylvania State University where he completed a BS in 2008, an MS in 2011 with focus on characterization of innovative textile-based energy absorbing devices, and a PhD in 2017 with a focus on design as a sequential decision process. His MS program was advised under Dr. Charles Bakis as part of the Vertical Lift Research Center of Excellence (VLRCOE) and worked closely with parallel efforts on fiber-matrix composite systems as well through the research team he managed. His PhD was co-advised under Dr. Timothy Simpson and Dr. Michael Yukish where they explored design decision-making from both a human-centric perspective and a computational one and, in particular, explored the implications of model-development (and their fidelity) on set-based design (SBD) iterative set-reduction. Miller's work at PSU ARL includes collaboration with a team of scientists and engineers. [swm154@arl.psu.edu]

Dr. Michael Yukish—is the Head of the Manufacturing Systems Division at Applied Research Laboratory of The Pennsylvania State University. He has been at ARL since 1993. The focus of his research at ARL is in the areas of multidisciplinary design optimization, simulation based design, and conceptual trade space exploration for complex systems. He also teaches courses in flight test and systems modeling with the Aerospace Engineering Department.. [may106@arl.psu.edu]

Abstract

There is a deficit of data on the detailed execution of design acquisition processes, data which is needed to truly understand and improve them. Simultaneously, the movement to digital engineering, and specifically model based engineering, offers a key opportunity to gather continual data needed to move acquisition processes forward. To address this issue, methods must be developed and implemented to capture key process metrics on the full product life cycle,

which includes conception, design, development, and test. The engineering acquisition process should be instrumented, capturing engineering metrics at a level of granularity sufficient to provide actionable information to other acquisition programs. These methods would be implemented on a set of diverse engineering programs, utilizing internal engineering design tools, product data and life-cycle management tools, and manpower reporting systems to capture data. This paper first discusses a number of specific examples of process instrumentation undertaken by the authors, then concludes with recommended lines of research for fully instrumenting acquisition processes.

Keywords: Digital Engineering ROI; Acquisition Process metrics; Systems Engineering Data Metrics; Design Process Improvements

Research Statement

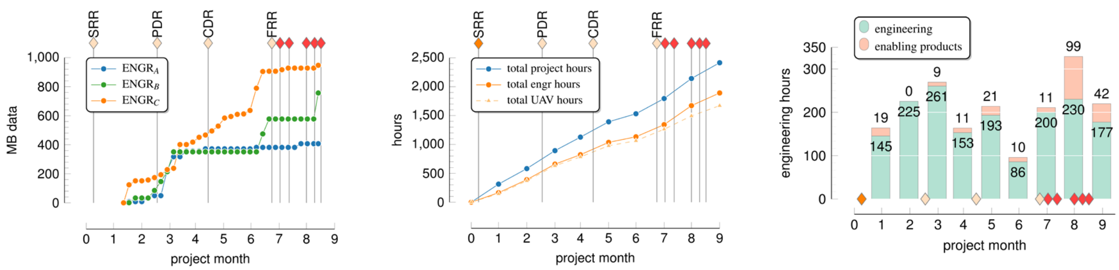
Fundamentally, the deep gathering of process metrics for large scale design efforts is not being done today. Processes must be measured and analyzed in order to improve. The three main roadblocks to gathering and sharing such data are (1) difficulties in data capture and categorization, (2) the proprietary nature of the data, and (3) a lack of emphasis from the community funding research on acquisition process improvement. Industry is historically unwilling and unable to provide design, schedule, and cost information at the level of granularity that can be analyzed to an actionable level. Cost, schedule, and design information provided by industry, per their contract requirements, is not at the level needed to decompose and correlate cost, schedule, and risk with design information. Additionally, there are minimal publications presenting metrics on design processes in the open literature, and those that do note the lack of data on which to base analyses.

Technical Approach and Relevant Background

As a Department of Defense (DoD) designated University Affiliated Research Center, ARL conducts essential research, development, and systems engineering in support of our nation's priorities free from conflict of interest with industry. We have had hundreds of engineering design and development programs covering a broad spectrum of engineering domains and varying in levels of complexity. For example, we have designed and built UAVs in months (Miller et al., n.d.), and we have designed and developed highly complex undersea systems in years (Penn State, 2022). We have over 1,000 full-time engineers and utilize a broad set of engineering application tools including requirements management databases, MBSE, CAD, engineering programming tools, product data and life-cycle management systems, and software management tools. We have implemented both traditional systems engineering approaches and highly adaptable, accelerated acquisition approaches that are more commonly used in fast prototype programs.

In addition to our design and development programs, we have been performing core research in Systems Engineering (SE) and design methods for over 20 years. Topics in this research include systems engineering, model based engineering, multidisciplinary design optimization, decision making, sequential decision making, Set Based Design (SBD), conceptual trade space exploration, data visualization, uncertainty propagation, and more (Martin & Simpson, 2004a, 2004b, 2006; Miller et al., 2013; Stump et al., 2002, 2004, 2009; Yukish et al., 2018).

Specific to process instrumentation, in a project funded by the Rapid Response Technology Office under the Office of the Secretary of Defense, ARL instrumented the design, development, and test efforts of a 3D-printed UAV, and captured process metrics on two full product life cycles (Miller et al., n.d.). In this effort, we tracked engineering hours to progress a UAV design from requirements to development and test, and published resulting data repositories for use in other data mining and systems engineering research. Results from this effort showed relationships between team performance and design iterations, and how rapid prototyping methods, made available by 3D printing, accelerated the design process. An example of process data collected and related to design review events is provided in Figure 1.



(a) cumulative digital data generated by engineers (summed weekly)

(b) project and tasking hours accumulated per month

(c) engineering and enabling technology hours at the start of each month

Figure 9: UAV Design: Engineering Data Metrics

Another example project involved data capture and analytics within the shipbuilding industry through the Institute for Manufacturing and Sustainment Technologies. The project analyzed Production Bill of Material (PBOM) data from a major shipbuilder and created a configurable web-based Artificial Intelligence (AI) based software application to enable more rapid robust and accurate error discovery. In the shipbuilding industry, undetected PBOM errors lead to re-planning and/or production rework, which results in significant cost increases. Current error detection and resolution involves manual examination of data by planners and is often done retroactively. To decrease errors and the labor-associated costs, ARL created a configurable, web-based AI software application, which enabled more rapid, robust, and accurate error discovery. This effort involved data collection, data mapping and mining, requirements development, current state process mapping, AI model prototyping and testing, and development of a final custom rules interface for more targeted PBOM error detection. The initial data mining and mapping efforts on this project were challenging. The database was highly connected, but not all of the connections were useful, and the process of algorithmically mining similarities would mistakenly relate fields or user-entered data. Figure 2 is an example of initial connectivity among the database tables analyzed.

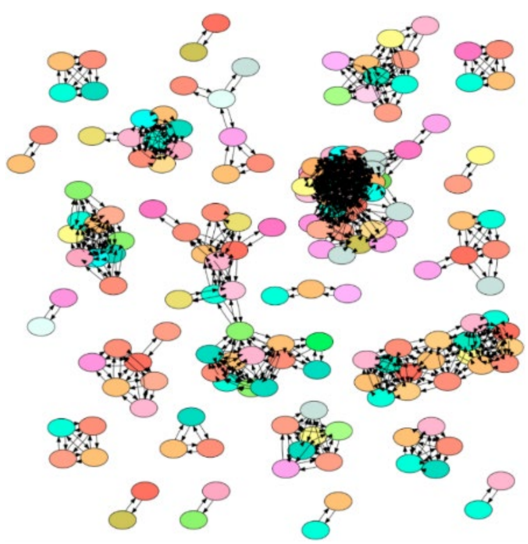


Figure 10: Example of Database Connectivity

Note: Colors Represent Different Tables, Nodes Represent Columns, Arrows Represent Inter-Table Relations

Several algorithmic approaches were developed and tested to automatically and correctly identify relationships among the data. One of the outcomes of this project was the development of a scalable and extensible mapping process by which complex data relations can be extracted from large data repositories to provide an automated platform for data analysis and the possible application of AI modeling techniques. These mapping processes directly enable applying AI techniques to analyzing extremely large datasets of design/manufacturing/acquisition-related data to make inferences on process performance, identify errors, and suggest actions for improvement.

ARL has been funded by the National Science foundation and Systems Engineering Research Center to research design decision support and SE design methods across a broad spectrum of topics. One topic focuses on treating design as a formal *Sequential Decision Process*, coupling Set Based Design (SBD) methods with Model Based Systems Engineering (MBSE), and developing new methods to progress through models of increasing levels of fidelity as the *consideration set* of possible solutions is narrowed (Miller et al., 2015, 2017; Yukish et al., 2018). In this research, we have developed qualitative and quantitative models of the decision process, and identified key data elements to capture in order to understand the state of an acquisition process. Conclusions from this research strengthen the need to better manage, maintain, and update design knowledge, and capture the state of the design in which decisions are made for later review and improvements (Yukish et al., 2018).

As a UARC, we are currently involved with several DoD program offices helping them define their digital engineering strategy, and their roadmap to generating digital twins of their assets so that they can reap the benefits of a digital twin in the design, test, and sustainment phases of their projects. Digital engineering and digital twinning both will create reams of data across the breadth of programs, from conception to sustainment that can be exploited to gain insight into the processes and identify areas for improvement. The data gathered for process *improvement*, however, is not necessarily the same as the data gathered to support the process itself. The time to identify and ensure process improvement data is up-front in the development of digital engineering and digital twin infrastructures.

Digital engineering and digital twin thrusts provide the perfect opportunity to instrument the design process, collect and analyze process data, and delivery methods to the acquisition community for future data analytics and process improvement. Industry partners that develop engineering systems for the DoD are a rich target for process data collection; however, these organizations are historically unwilling and unable to provide design, schedule, and cost information at the level of granularity that can be analyzed to an actionable level. ARL has performed multiple cost analysis studies for NAVSEA with the intent of better understanding cost and schedule risk areas in order to make acquisition improvements (Bennett, 2011; Clark & Bennett, 2008). However, the cost, schedule, and design information provided by industry, per their contract requirements, was not at the level needed to decompose and correlate cost, schedule, and risk with design information. As expected, this information, at lower levels of detail, is industry proprietary. As a university-affiliated research center with strong ties to the DoD acquisition community, we can instrument our processes to collect, categorize, group, and analyze engineering data from our programs, and provide analytical results and lessons learned to the acquisition community for future improvements.

Proposed Efforts and Expected Results

All organizations that strive to improve design processes will benefit from detailed, highly granular process data and the techniques to extract it from the Digital Engineering systems. Reducing defense acquisition cycle times has been a cornerstone goal of multiple DoD initiatives over the past several years. In addition, the OSD has published their Digital

Engineering Strategy, and requested all DoD components to develop digital engineering implementation plans that show desired outcomes. A crucial part of implementation success is showing improvement (reduced cost or schedules) from previous practice. To show improvement and to accelerate it, projects to support process instrumentation are needed.

Within the scope of these projects, the primary efforts should be to develop and implement data collection methods on engineering processes, update and improve methods based on initial data analysis, and provide these methods, analytics results, and data repositories to acquisition professionals, government labs, and industry. The projects should involve not just engineering design researchers, but also experts in business, contracts, legal, and logistics. Example relevant tasks for a research effort would include the following:

- Identify and categorize engineering tools and databases to be instrumented.
- Develop automated methods on these engineering applications to capture relevant data.
- Identify diverse engineering programs for initial design capture; collect and categorize historical data on the program and begin collecting data in real time. Choosing programs that have varying levels of design complexity and represent various engineering domains to assure methods are applicable across the larger engineering spectrum.
- Analyze data, identify data sensitivities and relationships, and modify methods as necessary.
- Analyze all data for statistical relationships and patterns.
- Provide data analysis results, lessons learned, data capture methods, and cleaned data repositories to Acquisition and Sustainment (A&S) for future implementation, study. and data mining.

In the initial stages, the acquisition instrumentation and analysis would be performed by dedicated researchers, as this is a nascent line of inquiry. In future phases of this work, the methods and processes should be transitioned to DoD acquisition programs and their industry performers to implement data collection methods *across all programs*, and implement developed process improvements, which will help reduce program schedule and costs.

The long-term objective of this line of research is to discover areas for acquisition engineering process improvements based on analysis of data on current and historical programs. The key tenet of the effort is **we cannot improve what we do not measure.** This tenet applies to every weapons system technology we acquire as they progress from basic research to developmental test, and on to operational test, at each stage rolling test results into improving the technology. This same dedication now needs to be applied to the acquisition processes themselves.

References

Bennett, L. (2011). *Studying the cost of Navy certification on manned submersibles.* Underwater Intervention Conference, February 22–24, 2011.

Clark, C., & Bennett, L. (2008). *Solving the complexities of submersible certification costs*. NDIA Joint Undersea Warfare Technology Fall Conference, September 8–11, 2008.

Martin, J. D., & Simpson, T. W. (2004a). *A Monte Carlo simulation of the Kriging model.* 10th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, AIAA-2004-4483, Albany, NY.

Martin, J. D., & Simpson, T. W. (2004b). *On using Kriging models as probabilistic models in design*. 2004 SAE World Congress, 2004-01-0439, Detroit, MI.

Martin J. D., & Simpson, T.W. (2006). A methodology to manage system-level uncertainty during conceptual design. *ASME Journal of Mechanical Design*, *128*(4), 959–968.

Martin, J. N. (1997). *Systems engineering guidebook: A process*.

Miller, S., Simpson, T., Yukish, M., Bennett, L., Lego, S., & Stump, G. (2013). *Preference construction, sequential decision making, and trade space exploration*. ASME 2013 IDETC/CIE 2013, Portland, OR.

Miller, S. W., Simpson, T. W., & Yukish, M. A. (2015). Design as a sequential process: A method for reducing design set space using models to bound objectives. In *ASME International Design Engineering Technical Conference & Computers and Information in Engineering Conference, DETC2015-46909, Boston, MA* (pp. 1–13).

Miller, S. W., Simpson, T. W., Yukish, & M. A. (2017). Two applications of design as a sequential decision process. In *ASME International Design Engineering Technical Conference & Computers and Information in Engineering Conference, DETC2017-68150, Cleveland, OH* (pp. 1–11).

Miller, S., Yukish, M., Hoskins, M., Bennett, L., & Little, E. (n.d.). *A retrospective analysis of system engineering data collection metrics for a 3D printed UAS design*. 17th Annual Conference on Systems Engineering Research (CSER).

Penn State. (2022). *Applied Research Laboratory, Undersea Systems Office*. https://www.arl.psu.edu/content/undersea-systems-office

Stump, G., Lego, S., Yukish, M., Simpson, T. W., & Donndelinger, J. A. (2009). Visual steering commands for trade space exploration: User-guided sampling with example. *ASME Journal of Computing & Information Science in Engineering*, *9*(4), 044501.

Stump, G. M., Yukish, M., Simpson, T. W., & Bennett, L. (2002). *Multidimensional visualization and its application to a design by shopping paradigm.* 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Atlanta, GA, AIAA.

Stump, G. M., Yukish, M. A., Simpson, T. W., O'Hara, J. J. (2004). Trade space exploration of satellite datasets using a design by shopping paradigm. In *IEEE Aerospace Conference, IEEE, New York* (pp. 3885–3895).

Yukish, M., Miller, S., Martin, J., Bennett, L., & Hoskins, M. (2018). Set-based design, model-based systems engineering, and sequential decision processes. *Naval Engineers Journal*, 130-4.

PANEL 26. GETTING AND USING THE RIGHT ACQUISITION DATA

Thursday, May 12, 2022	
1:55 p.m. – 3:10 p.m.	**Chair: Mark E. Krzysko,** Principal Deputy Director, Enterprise Information, Acquisition Data and Analytics ***Using Data Analytics to Detect Bridge Contracts*** Tim Hawkins, University of North Texas David I. Gill, Internal Revenue Service Jinsu Elhance, Data and Analytic Solutions, Inc. Robert Carlson, Data and Analytic Solutions, Inc. ***Making Federal Financial Data More Reliable With Emerging Tech*** David Gill, Internal Revenue Service Avram Ibrahim, ASI Government Chaudry Umer, Cooley LLP Sonia Jolly, Cooley LLP Alicia M. Miller, Internal Revenue Service ***Telling Time: Getting Relevant Data for Acquisition Schedule Estimating Relationships*** Charles Pickar, Naval Postgraduate School Brig Gen Raymond Franck, USAF (Ret.) ***Technical Data Package Independent Assessment*** Nicholas Haas, The MITRE Corporation Virginia Wydler, The MITRE Corporation Raymond Eresman, The MITRE Corporation Kathleen Hyatt, The MITRE Corporation Scott Patton, The MITRE Corporation Beverly Stark-Kublin, The MITRE Corporation

Mark E. Krzysko— is the Principal Deputy Director of Enterprise Information in the Acquisition Data and Analytics (ADA) organization. In this senior leadership role, Mr. Krzysko directs acquisition data governance, data access, and data science to enable the Department to make sound business decisions with data. He is leading a philosophical and technical transformation within the Department to make timely, authoritative acquisition information available to support insight and decision-making on the Department of Defense's major programs—a portfolio totaling approximately $2 trillion of investment funds over the lifecycle of the programs—as well as smaller programs and nontraditional acquisition approaches.

In 2016, Mr. Krzysko was invited to be a member of the White House Office of Science and Technology Policy's Data Cabinet, which was initiated to lead the effort to make leveraging the power of data the norm across the Federal government. In addition, he is a charter member of the Defense Science Interagency Working Group, which has a similar purpose.

Preceding his current position, Mr. Krzysko served as Assistant Deputy Under Secretary of Defense (ADUSD) for Business Transformation, providing strategic leadership for re-engineering the Department's business system investment decision-making processes. He also led efforts as ADUSD for Strategic Sourcing and Acquisition Processes and as Director of the Supply Chain Systems Transformation Directorate, championing innovative uses of information technologies to improve and streamline the supply chain process for the Department. As the focal point for supply chain systems, Mr. Krzysko led the transformation, implementation and oversight of enterprise capabilities for the acquisition, logistics and procurement communities.

In March of 2002, Mr. Krzysko joined the Defense Procurement & Acquisition Policy office as Deputy Director of e-Business. As the focal point for the Acquisition Domain, he was responsible for oversight and transformation of the acquisition community into a strategic business enterprise. This included driving the adoption of e-business practices across the Department, leading the move to modernize processes and systems, and managing the investment review process and portfolio of business systems.

From June 2000 to March 2002, Mr. Krzysko led the Electronic Commerce Solutions Directorate for the Naval Air Systems Command. From April 1991 until March 2000, Mr. Krzysko served in senior-level acquisition positions at the Naval Air Systems Command, including Contracting Officer of F/A-18 Foreign Military Sales, F/A-18 Developmental Programs, and the F-14. Mr. Krzysko began his career in the private sector in various executive positions including Assistant Managing Director for Lord & Taylor Department Stores and Operations Administrator for Woodward & Lothrop Department Stores.

Mr. Krzysko holds a Bachelor of Science Degree in Finance and a Master of General Administration, Financial Management, from the University of Maryland University College, and numerous certificates from Harvard University.

Using Data Analytics to Detect Bridge Contracts

David I. Gill—Procurement Data Scientist, Analytics, Research and Technology Division, Office of the Chief Procurement Officer, Internal Revenue Service. [david.gill@irs.gov]

Dr. Timothy G. Hawkins—Senior Subject Matter Expert, Procurement Data and Analytic Solutions, Inc. [Hawkins@DASconsultants.com]

Jinsu Elhance—Junior Data Scientist, Data and Analytic Solutions, Inc. [Elhance@DASconsultants.com]

Robert Carlson—Technical Manager/Lead Data Scientist, Data and Analytic Solutions, Inc. [RCarlson@dasconsultants.com]

Abstract

Bridge contracts—temporary contract actions that enable continued contractor performance until a replacement contract can be awarded—are not controlled and are suspected to be overused. While facilitating continued mission achievement, bridge contracts reduce competition, result in higher prices paid, and increase transaction costs. Yet, few agencies have a means to identify bridge contracts, meaning the extent of their use is unknown. Thus, most agencies do not identify, analyze, and monitor the risk associated with achieving statutory competition objectives. This research develops a data analytic methodology to identify bridge contracts, which can quantify the magnitude of the problem and serve as a starting point to enact policy to mitigate usage.

Keywords: Bridge Contract; Competition; Acquisition Planning

Disclaimer: The content of this article is the opinion of the writer and does not necessarily represent the position of the Internal Revenue Service.

Introduction

U.S. government contracts in 2020 accounted for over $682 billion (Bloomberg Government, n.d.). The U.S. government represents the single largest and unique business customer in the world; as such, it exerts an enormous economic importance (Boland & Godsell, 2021). Nevertheless, the business-to-government (B2G) market remains grossly understudied (Josephson et al., 2019). The public availability of data for the millions of contract actions annually has attracted recent scholarly attention of researchers and editors from top academic journals.

Although federal government agencies often have continuing needs for procured goods and services in order to meet mission needs, their contracts are time-bound. To meet agency needs, new contracts must be awarded in time to prevent a gap in coverage. However, given the long lead time for source selections, the complexity of the contracting process, budget uncertainties, an occasional lack of advance planning, personnel workloads, a lack of experience, and turnover, new contract awards are sometimes delayed (GAO, 2014; GAO, 2016). As a contingency, agencies utilize *bridge contracts*—"an extension to an existing contract beyond the period of performance (including option years), or a new, short-term contract awarded on a sole-source basis to an incumbent contractor to avoid a lapse in service caused by a delay in awarding a follow-on contract" (GAO, 2016, p. 4). Two defense agencies estimated their 2014 bridge contracts each exceeded $1 billion (GAO, 2016). Another study of the Department of Defense (DoD) identified 18 bridge contracts valued at $9 billion covering 2007–2011 (GAO, 2012). A study by the Institute for Defense Analysis estimated that over 23% of contracts reviewed were bridge contracts (Williams et al., 2012). A memorandum by the Undersecretary of Defense (AT&L) in 2018 reported 1,100 bridge contracts worth $13.7 billion awarded during 2015 (Longo, 2020).

While meeting mission needs, awarding bridge contract actions brings deleterious effects such as: reduced competition, potentially paying higher prices, and increased transaction costs (GAO, 2016). Such actions also delay or deny business opportunities to prospective suppliers. Despite these effects, currently, there is no mechanism to assess the frequency of the practice. Bridge contract actions are not reported in the Federal Procurement Data System-Next Generation (FPDS-NG; GAO, 2016). Since we do not know the frequency of bridge contracts, we also do not know the magnitude of the consequences or the antecedents (including their relative order of influence). Until the most prevalent causes are identified, agencies will be unable to manage and control the practice.

The purpose of this research, therefore, is to develop a data-analytic methodology—using natural language processing and graph network theory—to reliably identify bridge contract actions. Once applied, the model results will address two research questions.

(1) How prevalent is the practice of bridge contracts?

(2) Are there any discernable patterns in bridge contract use?

Based on the findings, the developed analysis methodology can serve as an internal control tool (per requirements of GAO/AIMD-00-21.3.1 [1999]) for agencies to mitigate overuse of bridge contracts. Findings could ultimately lead to policy changes that control the causal factors resulting in reduced usage. The DOD's use of "tripwires" for bridge contract length (Lucyshyn & Quist, 2019) and early policy by the Navy and Defense Logistics Agency resulted in reduced usage of bridge contracts (GAO, 2016). Consequently, competition rates should increase, opportunity should increase, and excessive prices and transaction costs should be avoided.

The remainder of this article is organized as follows. It begins with a review of the relevant literature surrounding competition, bridge contracts, and relevant theory. Next, the study presents the methodologies of quantitative data collection and analysis to explore the research questions. Lastly, discussions, limitations, implications, future research directions, and conclusions are offered.

Literature Review

Competition

Competition is the bedrock of federal acquisition (FAR 1.102(b)(1)(iii)). Perhaps chief among all of the benefits of competition is its instilled fairness—the ability to provide equal opportunity to all responsible suppliers (Doke, 1995). Absent fair opportunity, suppliers will cease pursuit, thereby decreasing competition (Doke, 1995). Competition also reduces the cost of procured goods and services, resulting in savings (GAO, 2016); increases return on investment for the taxpaying public; improves contractor performance; reduces fraud (GAO, 2012); and promotes innovation (Jackson & Alerding, 1997). Competition rates for the DoD in 2011 for all goods and services was 58%. The competition rate for non-research and development (R&D) service contracts was 78%, while the rate for R&D services was 59%. The competition rate for products was 41% (GAO, 2012).

Full and open competition is required by the Competition in Contracting Act (CICA) of 1984, with some exceptions. The most commonly cited exception for service contracts—excluding R&D—was "only one responsible source" (which includes unique capabilities, unsolicited proposals, particular follow-on contracts, intellectual property rights, standardization programs, and utilities). The second-most common exception was "authorized by statute" (e.g., Small Business Administration's 8(a) program, Federal Prison Industries, AbilityOne Program, HUBZONE Act, Veterans Benefits Act, and WOSB program). Other exceptions include: (1) an "unusual and compelling urgency" situation that does not afford adequate time to traverse the source selection process, (2) "industrial mobilization" to maintain a facility or manufacturer in the case of a national emergency, (3) "engineering, developmental, or research capability" to maintain essential services provided by education or non-profit institutions or by a federally funded research and development center, (4) "expert services" to support litigation or disputes, (5) "international agreement" for acquisitions reimbursed by a foreign government or wherein a treaty or agreement specifies or limits sources, (6) "national security" in cases in which disclosure of the need would compromise security, and (7) "public interest" (FAR 6.302). When the agency uses multiple-award indefinite delivery/indefinite quantity (IDIQ) contracts, orders must be competed unless an exception applies. Exceptions include: (1) urgency, (2) only one capable source, (3) economy and efficiency (including a logical follow-on order), (4) satisfying a guaranteed minimum dollar amount of the parent IDIQ award, (5) statutory requirement, and (6) small business programs. While there are several legitimate reasons to not compete requirements, some situations do not offer relief. There is no exception to competition available for situations wherein: (1) an agency fails to adequately plan for a contract action, (2) uncertainty about funds availability, or (3) expiring funds (e.g., at fiscal year-end). Notably, the exception due to a logical follow-on is used frequently—46% of all of the DoD's exceptions to fair opportunity cited this reason.

In addition to the situations identified as exceptions, other factors can affect competition such as program officials with preferences for particular contractors (e.g., incumbents), overly restrictive requirements, and unanticipated events such as bid protests (GAO, 2012). Unanticipated delays have also been attributed to program official turnover, requirements definition, expansion of requirements, late completion of pre-award documentation, workload, delays in source selection, an inexperienced workforce, budget uncertainties, and acquisition strategy approvals (GAO, 2014; GAO, 2016). These delays sometimes result in bridge contracts (GAO, 2012; GAO, 2016).

Bridge Contracts

A bridge contract is a useful tool to avoid a lapse in services (GAO, 2016), to preclude substantial duplication of costs, and avoid unacceptable delays (Jackson & Alerding, 1997). Bridge contracts are not a new phenomenon. Several comptroller general decisions of protests in the late 1980s prescribed their permissible usage under CICA (Jackson & Alerding, 1997), highlighting the care that agencies must exercise in using bridge contracts in order to avoid a CICA-based protest. Bridge contracts are used for a variety of services ranging from repair and maintenance of equipment, research and development, housekeeping, information technology and telecommunications, and professional administrative and management support, with the latter two accounting for half of bridge contracts (GAO, 2016).

The government awarded contracts worth $3 billion in Fiscal Year (FY) 2013 non-competitively on the basis of urgency (GAO, 2014). Noncompetitive contract actions using the urgency exception are limited in duration to: (1) the time needed to award a competitive contract, and (2) a maximum of one year unless a high-level approval is obtained (i.e., the head of the contracting activity or a designee). However, a study of DoD, Department of State, and the U.S. Agency for International Development revealed non-competitive contract awards based on urgency extending beyond one year in 10 of 34 contracts inspected—eight of which were later extended via modification to exceed the one-year limit (GAO, 2014). A separate study found six such contracts (out of 29 examined) that exceeded three years (GAO, 2016). Also noteworthy is that bridge contracts are sometimes (20/29 examined) succeeded by one or more additional bridge contracts (GAO, 2016) —one as many as five times increasing the projected value of the bridge by 264%.

The GAO (2014) found that justification and approval documents (J&A) did not always contain the required signatures, were sometimes ambiguous in supporting facts (e.g., the serious injury—financial or other—to the government) and were sometimes publicly posted to the FedBizOpps site late (beyond 30 days of award). These discrepancies can reduce the public's confidence in a fair and transparent contracting system—its fundamental goals (FAR 1.102(b)(3)).

As of 2016, agencies (DoD, Health and Human Services, and Justice) were unaware of the extent of their uses of bridge contracts (GAO, 2016). In fact, omitted from the FAR, agencies did not have a consensus definition of bridge contracts, and thus, most had no policy to manage and control their use.

Theoretical Foundation

The standard for determining that an agency did not adequately plan for an acquisition—an unallowable basis for an exception to competition—is low. "The GAO probably will find advance planning lacking only when there has been virtually no advance planning" (Jackson & Alerding, 1997, p. 218). Importantly, most of the causes of delays in sources selections—with proper planning—can be accommodated in acquisition milestones (e.g., review times, requirements documentation); however, some of them cannot, such as budget uncertainty, bid protests, and personnel turnover. "The practice of compliance is inherently a probabilistic activity due to administrative resource constraints, managerial error, misinterpretation and at times evasion" (Orozco, 2020, p. 258). Law and regulation will have varying degrees of uncertainty attributed to vagueness in language and weak regulatory enforcement (Orozco, 2020). Thus, a theory (or theories) explaining and predicting the use of bridge contracts must encompass: (1) a lack of adequate planning, and (2) uncertainty.

Agencies have taken advantage of the exceptions to competition. For example, the NSA used a bridge contract for installation and logistics services in 2013, allegedly to afford time to plan for a competition. Then, three weeks before its expiration—and thereby creating an urgent situation—the NSA awarded another sole source 8(a) contract (NSAIG, 2021). According to Marshall Doke, a prominent attorney in public contracting:

> Government agencies at least pay lip service to competition, but the actual users of supplies or services usually would prefer no competition at all and chafe at the rules and "red tape" of procurement procedures. The government users usually know the vendor they want or prefer, and describing their requirements adequately for competition in specifications or statements of work often is not a high priority. It is not surprising that specifications written around the product of a particular vendor are frequently developed. Nor is it surprising that government officials "split" a requirement to get below a specified dollar threshold for full competition.

A rule is "a statement of general applicability and future effect that implements, interprets, or prescribes law or policy or the organization, procedures, and standards for practice before an agency" (Rossi, 1995, p. 275). Rules help citizens and governments operate on the basis of common, predictable norms and conditions that will prevail (Rossi, 1995). However, rules cannot perfectly address all nuanced cases, nor can all values be formulated into rules. Thus, tension in law between rules and equity will persist. The principle of *regulatory equity* infuses equity into the regulatory process via an administrative exceptions process, which relieves a person or organization subject to a valid statutory or administrative rule from the legal obligation to comply. A challenge, then, is the case wherein the exception becomes the rule. At what point are exceptions to policy overused, and has this happened with respect to bridge contract actions?

Lindenberg's goal framing theory (GFT) serves as a basis for compliance theory (Etienne, 2011). Although more oriented toward the citizen regulatee, its principles can apply to bureaucrats who are charged with compliance. GFT posits that actors, in deciding whether to comply, pursue multiple different goals simultaneously. Three categories of goals are: hedonic goals (i.e., attaining pleasure or stimulation in task accomplishment), gain goals (i.e., to maintain or increase resources, commonly evaluated as cost versus benefits), and normative goals (i.e., to act appropriately). The three types of goals can operate in both noncompliant and compliant behaviors and can operate simultaneously. However, "for action to occur, one of these multiple goals takes precedence while the other goals take a secondary role, although not necessarily a negligible one" (Rossi, 1995, p. 306).

Per GFT, before taking action, a person evaluates the available options and prioritizes them according to their ability to attain goals. "Planned compliance or noncompliance epitomizes the intentional pursuit of various goals, such as to maximize one's utility, fulfill a moral obligation such as duty or trust, or dispose of one's fear of sanctions" (Rossi, 1995, p. 307). "Making a profit hinges on costs and benefits; it is consequentialist, whereas acting appropriately is not and instead hinges on whether the options available are congruent with internalized norms" (Rossi, 1995, p. 308). The most influential goal holds one's attention, referred to as the goal frame. Any other weaker goals are not top of mind.

In the context of public procurement, bureaucrats very likely trade off two goals—*gain* and *normative*. The gain goals pertain to organizational effectiveness. Whether a requiring activity is judged as effective may hinge on its ability to obtain and retain contractors supporting the mission. Source selections introduce risks of unknown contractors and risks of delays due to bid protests. They consume substantial human resources (Hawkins et al., 2016), effort that could otherwise be directed toward attaining organizational performance goals. While bureaucrats may know of the intent of regulation to promote competition, they may trade off normative compliance in order to gain on behalf of their organization, and, thus, their own performance appraisal in pursuit of job security or promotability. The exceptions afforded by FAR 6.302 give them an "out." Contracting communities, in the interest of customer support and mission impact, may be complicit in prioritizing gain goals (e.g., not forcing advance planning for a source selection or not denying customers contract coverage) over normative goals (e.g., promoting competition). In the absence of sanctions, rules violations may become ubiquitous because rule followers may feel cheated (hedonic goals) and that resources are wasted (gain goals), thereby suppressing normative goals (e.g., to do one's duty). In an experiment, visible evidence of [non]compliance influenced subjects to [not] comply (Keizer et al., 2008). Exposed to wall graffiti and a "no graffiti" sign, 68% littered. In the orderly situation (not exposed to graffiti), only 33% littered. This finding suggests that social acceptance can elevate utilitarian gain goals (e.g., avoiding a source selection) ahead of normative goals (e.g., promoting competition), a phenomenon not uncommon in the literature (Rossi, 1995).

Compliance has also been explored from the theoretical lenses of general deterrence theory (GDT; Malloy, 2003), which is grounded in classic criminology. "Classic Deterrence theory focuses on formal (legal) [and informal] sanctions and posits that the greater the perceived certainty, severity, and celerity (swiftness) of sanctions for an illicit act, the more individuals are deterred from that act" (D'Archy & Herath, 2011, p. 644). Nevertheless, to invoke GDT, the perpetrator must perceive the threat of a sanction. This may not be the case with bridge contracts. Alternatively, if GDT is relevant, the requiring activities and contracting personnel may not perceive a high certainty or severity of sanctions, if any. "Lack of enforcement is one general reason for failure of regulations to promote their underlying objectives" (Malloy, 2003, p. 453). Consistent rule enforcement is also important. According to Doke (1995, p. S-15), "In most cases, a bad rule is better than no rule, and consistent application and enforcement of a bad rule often is better than discretionary application and enforcement of a good rule."

A third theory potentially explaining bridge contract use is the *compliance norm*—that a firm, or as applied here, an agency, is a "law-abiding actor, struggling in good faith to comply with increasingly complicated and contradictory laws and regulations" rather than a pure profit maximizer (Malloy, 2003, p. 454). An extension of this theory considers firm compliance as vulnerable to its organizational routines. Here, the firm is "a system for allocating and managing resources necessary to achieve [its] goals" (Malloy, 2003, p. 458). "In this 'systems' view of the firm, noncompliance is seen as a management problem: it is the complexity of the firm, rather than just the complexity of the regulations, that lies at the root of regulatory violations" (Malloy, 2003, p. 459). Organizations are plagued by their complex functional organizational structures that disperse information, resources, and accountability. An example in a government context is the functional separation of contracting, requiring activities, and accounting and finance. The contracting organization might "own" responsibility for competition compliance, while requiring activities may not have competition incorporated into their processes. These "deficient routines" prevent an organization from being able to comply. Without adequate processes, requiring activities succumb to more uncertainty (i.e., unplanned events such as understaffing, turnover, lack of training, and changing budgets) that cause a need for contract extensions.

Methodology

This research relied upon multiple data analysis methodologies. We use natural language processing (NLP) to prepare the data. Then we apply two machine learning algorithms via random forest regression to predict likely bridge actions—one for bridge modifications and one for bridge contracts. The remainder of this section explains assumptions made, the data sources, data preparation, and model construction.

Advanced analytics techniques are helpful, but not perfect, in detecting bridge contracts. Network analysis is a branch of integrated analysis techniques that allows one to explore relationships between interconnected agents or parties. A network is an operative analytical tool that applies the mathematical language of graph theory and linear algebra and outlines connections. This approach provides insights into the nature of relationships that a particular agency has with specific vendors. In each network, like the one shown in Table 1, each party/agent/person is represented as a node and their interactions amongst each other are represented as edges. Note that an interaction can be any point of relation, like a recipient of a contract or a type of contract and the vendors that can fulfill it. Additionally, fuzzy matching principles are important in connecting bridges with predecessors. Exact or approximate matches in some, but not all, relevant data fields may be sufficient to identify bridges and predecessors. Variation was observed in the timing of original contract end dates and bridge start dates. Other relevant data fields varied occasionally too (e.g., new PSC code selected, new Funding Office codes established).

Table 1. Example of a matching relationship with some matching and non-matching fields (Source: FPDS.gov)

Field	Original Contract (Predecessor)	Bridge Contract
PIID	TIRNO11D000160019	2032H518F00765
POP Start Date	09/30/2016	09/26/2018
Ultimate Completion Date	09/29/2018	09/26/2019
Contract Length	2 Years	1 Year
Contract Value	$35,787,420.44	$27,555,770.14
Funding Office Name	Commissioner	Commissioner
Vendor UEI	CKV2L9GZKJK3	CKV2L9GZKJK3
Number of Offers	2	1
Product/Service Code Description	SUPPORT- PROFESSIONAL: PROGRAM EVALUATION /REVIEW/DEVELOPMENT	SUPPORT- PROFESSIONAL: OPERATIONS RESEARCH/ QUANTITATIVE ANALYSIS
Award Description	OCA Compliance and Data Analytics Support Services	Compliance and Data Analytics Support Services
Extent Competed	Full and Open Competition	Full and Open Competition (*miscoded)
Fair Opportunity / Limited Sources	Fair Opportunity Given	Urgency

IRS contract documents and a FY2021 Navy bridge contract report were used as a source of ground truth to assess the model. Model precision, the proportion of algorithmically identified putative bridge contracts that are actually bridges was calculated. The formula for precision is True Positives / (True Positives + False Positives).

Methodology Assumptions

At the inception of this research, we outlined a list of assumptions about bridge contracts, bridge modifications, and the data, which were used as a basis in the development of the algorithm.

1. A bridge contract will have the same vendor name, funding office, and department as the contract it is bridging (i.e., its predecessor).
2. Procurement instrument identification numbers (PIID) —contract numbers—were considered erroneous if less than four characters.
3. A bridge modification will share the same award PIID as its modified award.
4. Service contracts can be unilaterally extended for up to six months via modification under the authority of FAR 52.217-8. The scope of this research includes extensions under the authority of FAR 52.217-8, which is consistent with the GAO's definition of bridge contracts and is a common mechanism used (GAO, 2016). Notably, not everyone agrees that a bridge contract encompasses an extension under FAR 52.217-8 (Longo, 2020).

Data Source and Preparation

Contract Award Data was downloaded from USASpending.gov and used to develop and train the model. The Navy provided a report of bridge contracts and modifications from FY2021 which was used to classify entries in the data. Additionally, IRS procurement documents and standard forms (Standard Form [SF] 1012 Limited Sources Justification, SF 1013 Justification for Other Than Full and Open Competition (JOFOC), SF 1014 Justification for an Exception to

Fair Opportunity (JEFO)) were scanned using NLP techniques to identify bridge keywords, providing additional labels for the data. A total of more than five million contract actions were compiled for analysis covering the period from FY2010 to FY2022.

Both the Bridge Contract Classifier and Bridge Modification Classifier were framed as problems within the conceptual framework of graph theory. Essentially, a graph can be thought of as a collection of objects called "vertices" and relationships between those objects, called "edges." Bridge actions, whether contracts or modifications, are heavily context-dependent occurrences. Consequently, creating a network that captures how data entries are related to one another offers an advantage in the task of identifying anomalies. Both classifiers made use of a "directed" graph structure, wherein edges can be thought of as arrows, linking an award A to an award B but not award B to award A. Otherwise, the structure of the network differed slightly between each classifier to better suit the classification task.

Certain column values needed to be transformed to be fed into our random forest models. Product or service codes were categorized into one-hot encodings which indicated whether the award was for knowledge-based services, logistics or transportation services, operations services, or medical services/supplies.

Bridge Contract Classifier Methodology

As received from the USASpending.gov website, the contract data to be used in the Bridge Contract Classifier needed to be processed. The first step was to aggregate properties of awards over their full life cycles. This was done by chronologically ordering the base award and modifications of a shared PIID and aggregating values such as period of performance and contract value. Any contracts with a PIID of less than 4 characters were seen as erroneously entered and were consequently removed. After this process of simplifying our data and trimming out errors, we were left with a data population (*n*) of 954,218.

To place these data entries into our graph, we represented contracts as vertices, and defined a set of rules governing whether or not an edge should be drawn between them. Creating a set of constraints on which edges can exist prevents the graph from becoming too complex. With fewer edges, there is less computation required to assess whether an edge may be linking a bridge award to its predecessor. The constraints dictated that an edge should only be directed from a potential predecessor to a potential bridge if the two awards shared a vendor, shared a funding office, and were chronologically sequential. The chronological ordering prevented the construction of edges pointing "backwards in time," as a bridge will never precede its predecessor. Understanding that contract award dates can differ from actual performance dates, we also remove any edges where the period of performance gap is less than seven, meaning the potential bridge began more than seven days before the potential predecessor ended.

The computational complexity of creating a graph which represented chronologically-ordered edges between all awards with the same vendor and funding office required that we train our model on a subset of the data. To do this, we sampled 1,000 vendor-funding office "clusters" from our graph and ensured that the clusters containing our labelled entries were included. This resulted in a final data set size of awards (vertices) = 4,904 and edges = 82,645. Those contract actions flagged as bridge contracts can be verified with additional data from justifications and approvals (J&A) available on SAM.gov or from agency reports that track bridge actions.

We create features using our data that we believe will provide signals for classifying bridges. Plots and descriptions of these features are provided in Figure 1. The first feature we created was similarity of award description. After tokenizing and stemming the award description field of the individual awards, we scored the similarity of descriptions across edges in our graph. Figure 1 shows that bridge edges likely do not share a description similarity distribution with non-bridge edges.

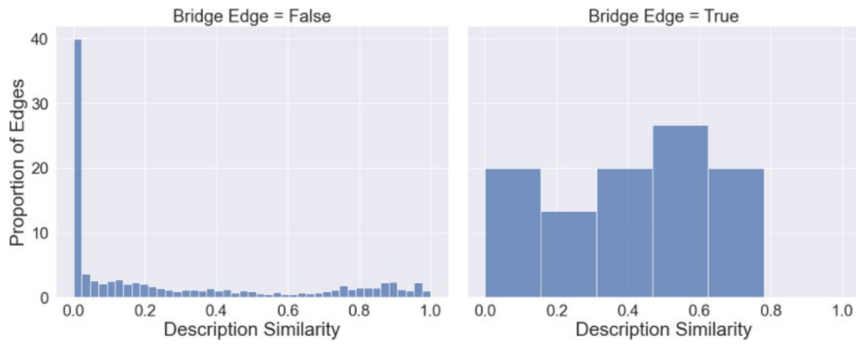


Figure 1. Comparing Award Description Similarity for Bridge and Non-bridge Edges

To follow up on an initial assumption that bridge awards have a similar value per day to their predecessors, we created a feature that flags whether or not the percent change in value per day is greater than 50% and less than 300%. Figure 2 shows the distribution of bridges and non-bridges with similar values per day.

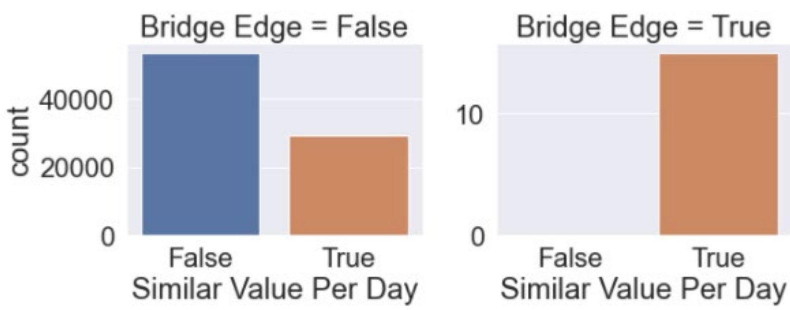


Figure 2. Comparing Edges where 0.5 < Percent Change in Value per Day < 3 for Bridge and Non-bridge Edges

Bridge contracts are typically used as a vehicle for covering a temporary lapse in service at the end of an expiring contract. To capture this, we calculated the number of days between the end of a predecessor's period of performance (PoP) and the beginning of a potential bridge. This is shown in Figure 3.

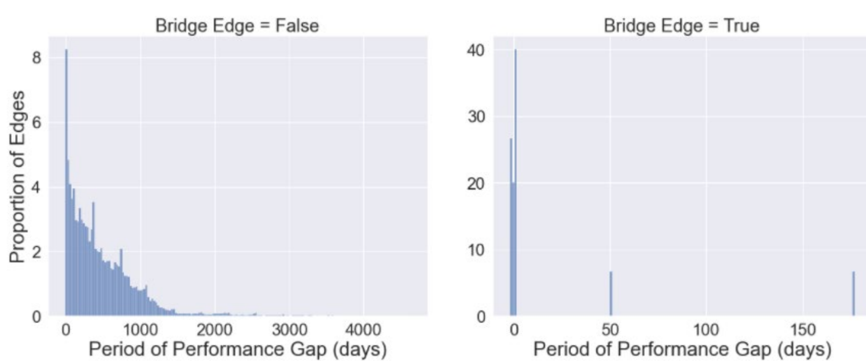


Figure 3. Comparing PoP Gap for Bridge and Non-Bridge Edges

Due to the irregular circumstances presented by the COVID-19 pandemic, funding was specifically allocated to certain awards to maintain service. We have flagged those as shown in Figure 4.

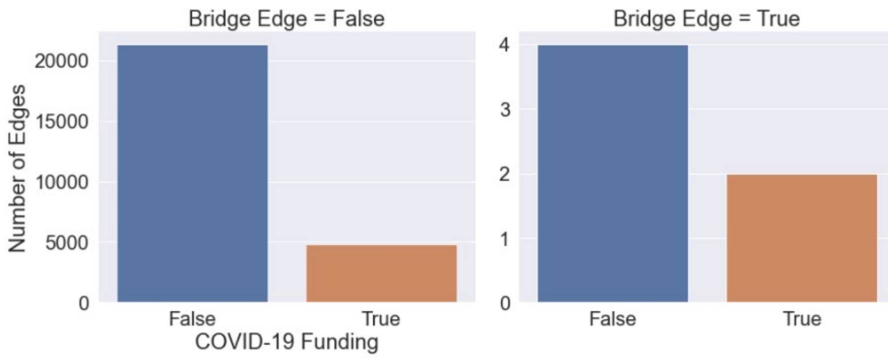


Figure 4. Number of Edges Where Either Award Received Funding Related to COVID-19 (only counts edges where the predecessor's period of performance began in 2019 or later)

The end of the fiscal year could also bring urgency to procurement that encourages the use of bridge contracts. We created a flag for edges where the predecessor expires within the 4th quarter of the fiscal year.

Figure 5. Shows Count of Edges Where Predecessor Expires in 4th Quarter of Fiscal Year

Bridge contracts do not follow proper procedure for allowing potential vendors to compete for the award. As a result, the fair opportunity sourcing code field in USASpending.gov provides an important indicator of how vendor sourcing took place. Table 2 shows the counts for edges wherein the potential bridge was flagged as a "follow on action following competitive initial action." After these features were created, the data was ready for the Random Forest Classifier, which has 50 decision trees.

Table 2. Edge Counts for Bridge Categories and a Fair Opportunity Sourcing Category of Interest

	Bridge Edge	Non-Bridge Edge
Follow-on Action after Initial Competition	2	163
Other Fair Opportunity Sourcing Code	13	82467

Bridge Modification Classifier Methodology

The same data set was used for the development of the bridge modification classifier. The graph was created to represent modifications as the vertices with edges pointing backwards in time such that the second modification on an award points to the first modification on the same award, and that first modification points to the base award (mod number = 0). We were able to process a much larger sample of data as the clusters of modifications were significantly smaller due to the fact that edges only existed between modifications that shared an award PIID. Without needing to sample our data as heavily as with our bridge contract data, we were left with vertices = 626,354 and edges = 38,279. The differences in vertices and edges is largely accounted for by awards which have no modifications. This allowed for additional data fields to be defined as shown in Table 3.

Table 3. Data Features for Modification Edges

Feature	Definition
Action Rank	Modification order as determined by action date
Action Value	Change in total dollars obligated
Action Performance Days	Period of performance of a modification (i.e., 1-month extension)
Value Per Day	Total dollars obligated divided by number of days in period of performance
Exercise Rank	The number of times an option has been exercised on an award
Number of Options	An approximation of the number of options available to the award; calculated as the number of years in the period of performance of the initial contract award
Potential End Date Delta	Period of performance extension caused by modification; calculated as difference in period of performance potential end dates

At this point, we have the training data necessary for our model. The model trained is a Random Forest Classifier (as defined in the sklearn.ensemble). We maintain the default parameters for the model and train a set of 50 decision trees.

Results

Bridge Contract Classifier Results

Figure 6 shows the mean decrease in impurity (MDI) for our model for each feature included in the training set. The model was trained and run 100 times and the feature importance plot has taken the mean MDI across those iterations. Features suffixed by an "_x" belong to the potential bridge contract while features suffixed by "_y" belong to the predecessor.

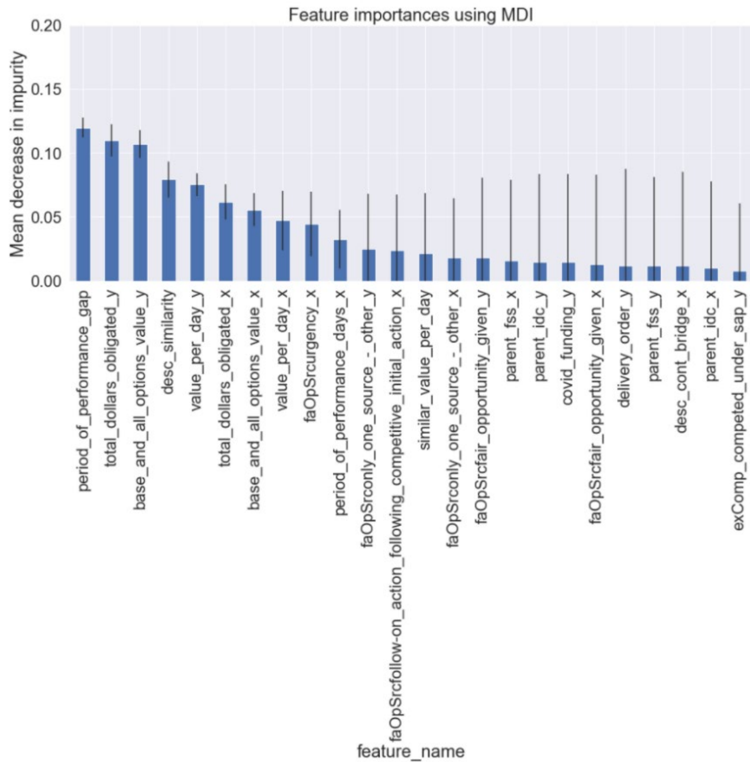


Figure 6. Mean Decrease in Impurity for Top 24 Features Across 100 Iterations

The outputs of our model are dependent upon the threshold we set for its predictions. By default, the model will only classify something as a bridge if more than half of the decision trees agree that it should be labelled as such. This produces Figure 7. The confusion matrix shown in Figure 7 shows that the model produced 13 true positives and 2 false negatives. The precision of the classifier is defined as the number of true positives divided by the total number of predicted positives. Precision can be seen as the accuracy of the model's positive outputs. The recall is calculated as the number of true positives divided by the total number of positives in the test data and describes the probability that a bridge is correctly identified.

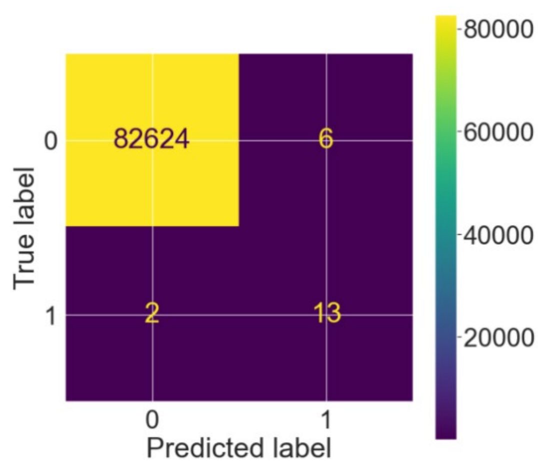


Figure 7. Confusion Matrix From the Random Forest Model's Output Precision=.684, Recall=.868

Bridge <u>Modification</u> Classifier Results

With fewer labelled bridge modifications than bridge contracts, the outputted model ran the risk of becoming highly biased toward the training set. However, the outputs of the model's feature importance align closely with our initial assumptions for important signals of bridge modifications. Figure 8 depicts the feature importance breakdown for the bridge modification random forest. For bridge modification features, "_y" indicates a data point belonging to a preceding modification and "_x" fields belong to the potential bridge mod.

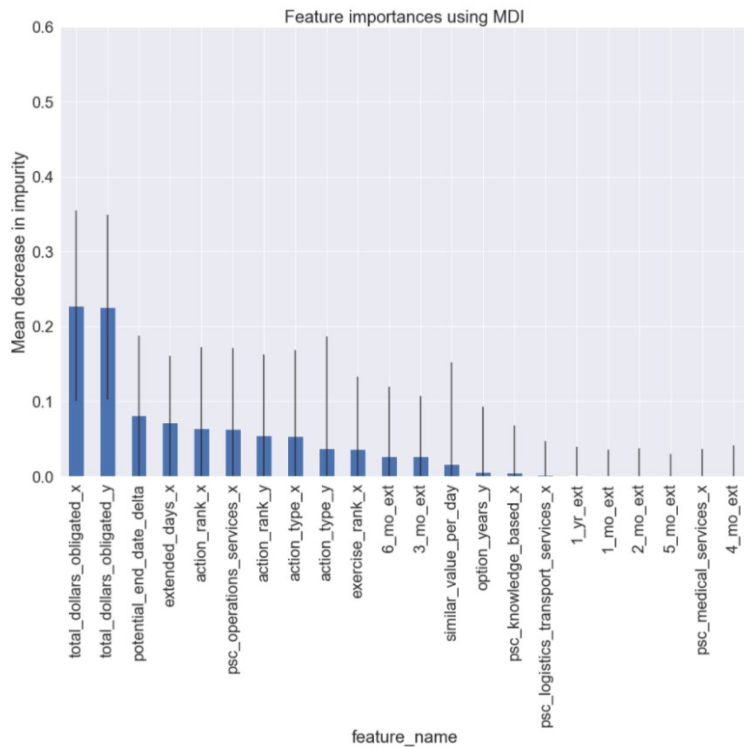


Figure 8. Mean Decrease in Impurity for Top 24 Features Across 100 Iterations

As can be seen in the confusion matrix shown in Figure 9, our model has produced 2 true positives and 2 false negatives. The model's precision indicates that roughly 66% of its positive outputs will be bridges; its recall indicates that there is a 50% chance that a bridge is correctly identified.

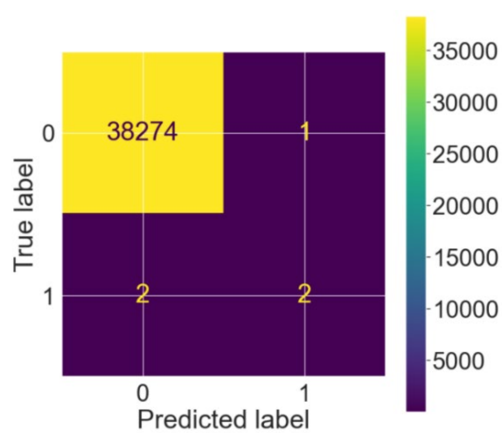


Figure 9. Confusion Matrix From the Random Forest Model's Output Precision = .667, Recall = .5

Results Conclusions

The results of both models are highly biased due to the few labelled bridges that we provided in training. As a result, the models tend to be extremely conservative, claiming that an entry is a bridge only if it has seen a very similar combination of features being labelled as a bridge in the past. However, in their current state, these models output a few potential bridges which are not labelled as bridges but seem to fit the pattern of bridges in the training set (i.e., false positives). To continue to improve the models, these cases should be investigated, and if determined to be bridges, fed back into the models as training data.

Identifying which awards were either bridges or contained a bridge modification within them can be done by closely investigating the model outputs. Each model outputs a table of actions (contracts or modifications) and the model's predicted "probability of being a bridge." Revising these outputs and confirming the model's predictions will be key to the longevity and continued improvement of this analysis. At the moment, the model is already turning up contracts and modifications which appear to be bridges upon closer inspection, but were not labelled as such in the training data.

Assumption Conclusions

Some of our initial assumptions did not hold true. Our assumption that vendor name and contract NAICS code would remain the same between predecessor and bridge contract applied in a majority of cases, but not always (example 1: vendor legally changed name; example 2: a bridge contract was given to a new vendor while a contract award was under protest). However, this was identified to be an outlier case, the final bridge contract edge constraint requires vendor name and NAICS code equivalence. With greater computing power, this edge constraint could be loosened, capturing these edge cases.

Discussion

The purpose of this research was to develop a data-analytic methodology to reliably identify bridge contract actions. The following two research questions were addressed.

(1) How prevalent is the practice of bridge contracts?

(2) Are there any discernable patterns in bridge contract use?

Previous efforts to identify the prevalence of bridge contracts have suggested that the proportion of actions which are bridge actions is greater than the proportion found by our models. This is very likely due to the lack of labelled bridge contracts and modifications in the training data. For example, in our bridge contract model outputs, many of the awards with non-zero probabilities of being bridges actually contained the word "bridge" in their award descriptions. This shows that our model is likely turning up only the most obvious bridge contracts, and those which do not "self-identify" likely slip through the cracks—evidence of a biased model. Consequently, the prevalence of bridge actions may be significantly greater than our model suggests.

As for research question number 2, the most predictive features of bridge contracts include (in order of predictive power): the period of performance gap between the predecessor and bridge contract action, the base and all options value, dollars obligated, similarity in the award descriptions, a consistent value per day among the predecessor and bridge contract action, urgency, period of performance days, competition codes in FPDS, and COVID-19 funding. The most predictive features of bridge modifications include (in order of predictive power): dollars obligated, potential end date, action rank, services, action type, exercise rank, six or three month extension, similar value per day, option year, knowledge-based services, and logistics and transportation services.

Managerial Implications

The GAO routinely cites its “Standards for Internal Control in the Federal Government” (GAO, 1999) to determine that agencies do not “identify, analyze, and monitor risks associated with achieving objectives, and that information needs to be recorded and communicated to management so as to achieve agency objectives” (GAO, 2016, p. 10). Since NLP is instrumental in linking contract action data from USASpending.gov to procurement documents, those documents (e.g., sole source justifications and approvals [J&A], memoranda for record, price negotiation memoranda) should explicitly identify bridge actions. Additionally, since bridge actions are commonly succeeded by another bridge action, documents in the contract file (e.g., sole source J&As) should report the history of prior extensions and identify the original contract extended. Together, this data would enable agencies to efficiently create dashboards to properly track and manage their use of these actions.

Procurement activities should consider using the modelling process developed herein to analyze their contract data to identify potential bridge contract actions. This will require either routinely contracting for the analyses or developing an organic data science capability including natural language processing and machine learning.

Study Limitations

This research is not without limitations. First, federal procurement data in FPDS-NG is riddled with errors. For example, coding of non-competitive contract actions due to urgency in FPDS-NG was found to be grossly erroneous in 2014—45% miscoded (GAO, 2014). Thus, any analyses based on this data is likely faulty to some degree.

We do not have an understanding of the true prevalence of bridge modifications. As a result, no assertions can be made that we have found a representative sample of labelled data for our model to be trained on. This results in a model which is highly biased to very few flagged data points. As more data becomes available on true bridge modifications, the model can be improved through iterative training.

Future Research Directions

Future research could refine the prediction models by adding features to improve predictive accuracy. While omitted from the scope of this study, future research could explore analytical techniques to detect edges between successive bridge actions. This is important since the GAO’s analysis (2016) revealed five requirements that were bridged multiple times each (from 3 to 12 times). Future research could apply the methodology developed herein to contract actions of government organizations beyond the IRS and U.S. Navy.

Conclusion

Bridge contract actions pose problems such as impeding competition, paying higher prices, and increased transaction costs. Few agencies rigorously identify and control bridge contracts. This research developed a data analytic methodology to identify bridge contracts. The most predictive features of bridge contracts and bridge modifications include: the period of performance gap between the predecessor and bridge contract action, the base and all options value, dollars obligated, similarity in the award descriptions, a consistent value per day among the predecessor and bridge contract action, urgency, period of performance days, competition codes in FPDS, COVID-19 funding, potential end date, action rank, services, action type, exercise rank, six- or three-month extension, option year, knowledge-based services, and logistics and transportation services. Advanced analytics techniques are helpful, but not perfect, in detecting bridge contracts. With continued development and iterative training the models have the potential to produce more accurate results.

Acknowledgement

Special thanks to Mrs. DeCarol Jovanovic, ASN (RDA), DASN (P), for assistance with bridge contract data collection.

References

Boland, M., & Godsell, D. (2021). Bureaucratic discretion and contracting outcomes. *Accounting, Organizations and Society*, 88, 1–24.
D'Arcy, J., & Herath, T. (2011). A review and analysis of deterrence theory in the IS security literature: Making sense of the disparate findings. *European Journal of Information Systems, 20,* 643–658.
Doke, M. J., Jr. (1995). Competition requirements in public contracting: The myth of full and open competition. *Federal Contracts Report*. *64*(3), S3-S38. Bureau of National Affairs.
Etienne, J. (2011). Compliance theory: A goal framing approach. *Law & Policy, 33*(3), 305–333.
Federal Acquisition Regulation (FAR).
FAR, 48 C.F.R 1.102(b) (n.d.).
FAR, 48 C.F.R 6.302 (n.d.).
FAR, 48 C.F.R 52.217-8 (n.d.).
GAO. (1999). *Standards for internal control in the federal government* (GAO/AIMD-00-21.3.1). https://www.gao.gov/assets/files.gao.gov/assets/aimd-00-21.3.1.pdf
GAO. (2012). *Competition for services and recent initiatives to increase competitive procurements* (GAO-12-384). https://www.gao.gov/products/gao-12-384
GAO. (2014). *Noncompetitive contracts based on urgency need additional oversight* (GAO-14-304). https://www.gao.gov/products/gao-14-304
GAO. (2016). *Defining and tracking bridge contracts would help agencies manage their use* (GAO-16-15). https://www.gao.gov/products/gao-16-15
Hawkins, T., Gravier, M., & Yoder, E. C. (2016). Federal bid protests: Is the tail wagging the dog? *Journal of Public Procurement, 16*(2), 152–190.
Jackson, A., & Alerding, S. A. (1997). Expanding contracting opportunities without competition. *Public Contract Law Journal, 26*(2), 205–220.
Josephson, B. W., Lee, J. Y., Mariadoss, B. J., & Johnson, J. L. (2019). Uncle Sam rising: Performance implications of business-to-government relationships. *Journal of Marketing, 83*(1), 51–72.
Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *ScienceXpress, 322*(5908), 1681–1685.
Longo, D. P. (2020). On contracting: Avoid bridges to nowhere. *Army AL&T Magazine*, (Summer). https://asc.army.mil/web/publications/army-alt-magazine/
Lucyshyn, W., & Quist, S. (2019). Evaluating the impact of contracting "tripwires" on service acquisitions (UMD-19-200). University of Maryland Center for Public Policy and Private Enterprise.
Malloy, T. F. (2003). Regulation, compliance, and the firm. *Temple Law Review, 76*(3), 451–531.
National Security Agency Office of the Inspector General. (2021). *Audit of National Security Agency's facilities and logistics service contract* (AU-18-0015). National Security Agency.
Orozco, D. (2020). A systems theory of compliance law. *University of Pennsylvania Journal of Business Law, 22*(2), 244–302.
Rossi, J. (1995). Making policy through the waiver of regulations at the Federal Energy Regulatory Commission. *Administrative Law Review, 47,* 255–301.
Williams, L. M., Rose, S. L, & Rehwinkel, A. S. (2012). Competitiveness in the services sector: Understanding the contracting data. D-4659. Institute For Defense Analyses.

Making Federal Financial Data More Reliable With Emerging Tech

David I. Gill—Procurement Data Scientist, Analytics, Research and Technology Division, Office of the Chief Procurement Officer, Internal Revenue Service. [david.gill@irs.gov]

Avram Ibrahim, MBA—Program Manager, ASI Government. [aibrahim@ASIGovt.com]

Chaudry Umer, Esq.—Attorney, Cooley LLP. [uchaudhry@cooley.com]

Sonia Jolly—Accountant, Office of the Chief Financial Officer, Internal Revenue Service. [sonia.jolly@irs.gov]

Alicia M. Miller—Director, Analytics, Research and Technology Division, Office of the Chief Procurement Officer, Internal Revenue Service. [alicia.m.miller@irs.gov]

Abstract

Federal agencies are stewards of billions in taxpayer funds. Given the scale of federal financial transactions, maintaining reliable, high-quality financial data can be challenging. The use of emerging technologies such as robotic process automation (RPA) and natural language processing can reduce manual work for agency employees and improve the consistency of financial data. These technologies are key to success on financial audits and maintaining public confidence in the reliability of procurement and nonprocurement financial information.

Legislative Goals and Historical Perspective on Federal Contract Spending

Federal government spending has not always been open to public scrutiny. In fact, public access to government spending is a recent development. Prior to World War II, the key federal law controlling disclosure of government information was the Housekeeping Statute of 1789 (now codified at 5 U.S.C. § 301). This statute gave the federal government the "general authority to operate their agencies" and withhold records from the public. Information restrictions continued to prevail in the United States during the first half of the 20th century, as federal agencies claimed exemptions from disclosing such data due to security risks associated with pre- and post-wartime activities (Yu & Robinson, 2012). As such, the public remained unaware of how the government spent federal dollars.

The passage of the Freedom of Information Act (FOIA) in 1966 was a breakthrough for advocates of a more open and transparent government. Through a FOIA request, any person now has a right, enforceable in court, to obtain access to federal agency records. The history of the FOIA is important because the act continues to serve as the foundation for all transparency-related initiatives. The FOIA continues to evolve and respond to the changes in technology as it fulfills its objective of providing access to government information through electronic media.

In 1980, a report from the U.S. Comptroller General described the newly created Federal Procurement Data System (FPDS; General Accounting Office, 1980). The system was created to collect government-wide information on what is bought, who bought it, when, where, and how the contract was awarded. FPDS was created to serve several purposes:

- Enable Congress to make informed public policy decisions relating to procurement programs.
- Provide the executive branch with information necessary for managing the procurement process.
- Support interagency acquisition activities.
- Furnish suppliers with information on federal agency needs and enhance competition. (General Accounting Office, 1980)

The report noted some challenges with collecting data from agencies in a timely manner and ensuring the accuracy/completeness of reporting. Nonetheless, establishing FPDS was a significant milestone in collecting standardized, electronic data on federal procurement purchases.

On April 6, 2006, Senator Tom Coburn and Senator Barack Obama introduced S. 2590 in the Senate (Federal Funding Accountability and Transparency Act, 2006). The bill proposed a more proactive approach to transparency, where the federal government would connect disparate data sets to provide a more comprehensive picture of federal contracts and grants spending. Anyone with access to the internet would be able to download transaction-level details related to federal grants and contracts. This accessibility would eliminate the various inefficiencies and hurdles resulting from formal FOIA requests for such data.

Congress has enacted a number of statutes regarding contracting and other financial information. The Federal Funding Accountability and Transparency Act (FFATA) of 2006 requires "disclosing direct Federal agency expenditures and linking Federal contract, loan, and grant spending information to programs of Federal agencies to enable taxpayers and policy makers to track Federal spending more effectively." Further, the data verification and validation burden on federal employees can be eased thru RPA. The FFATA mandates "consistent" and "reliable" contract spending data. At the bill's signing, President Bush noted that the government issues more than $400 billion in grants and more than $300 billion in contracts annually. "Taxpayers have a right to know where that money is going, and you have a right to know whether or not you're getting value for your money," the president said. "By allowing Americans to Google their tax dollars, this new law will help taxpayers demand greater fiscal discipline" (Government Contractor, 2006).

The Digital Accountability and Transparency Act of 2014 (DATA Act) was enacted to

- Expand the Federal Funding Accountability and Transparency Act of 2006 (31 U.S.C. 6101 note) by disclosing direct Federal agency expenditures and linking Federal contract, loan, and grant spending information to programs of Federal agencies to enable taxpayers and policy makers to track Federal spending more effectively;
- Establish Government-wide data standards for financial data and provide consistent, reliable, and searchable Governmentwide spending data that is displayed accurately for taxpayers and policy makers on USASpending.gov … ;
- Improve the quality of data submitted to USASpending.gov by holding Federal agencies accountable for the completeness and accuracy of the data submitted. (Digital Accountability and Transparency Act, 2014)

The U.S. government is a global leader in spending transparency. A 2017 study identified the United States as one of several "advanced jurisdictions" with respect to open data (Maurushat et al., 2017). Foreign governments such as the United Kingdom, Netherlands, and Australia have been inspired by U.S. open spending data policies (Huijboom, 2011). Another article cited the U.S. DATA Act as an inspiration for Dubai's open data law (Rizvi, 2016).

Extramural Acquisition Research and Journalism

A literature review was conducted identifying research and news stories based on open federal spending data. A high level of data quality (timeliness, accuracy, completeness) is important to research integrity and public confidence. Ready availability and improving data quality and completeness has spurred extramural research involving federal spending. Data portals such as USASpending.gov have enabled research studies in a variety of professional fields. A paper by university researchers explains, "The opening of datasets in machine readable linked data is of particular importance to university and private industry researchers as it opens hundreds of thousands of previously private datasets to be used for new research" (Maurushat et al., 2017). Data from USASpending.gov is available in a granular, disaggregated formation with detailed information each grant, contract, modification to a contract, and so forth.

Table 1. Examples of Extramural Research with Open Federal Spending Data

Professional Field	**Organization**	**Study**
Global Development	Columbia University	U.S. Spending in Haiti
Public Affairs	University of Missouri	Federal Contracting Trends in Missouri
Business Innovation	Hoover Institution	Supporting Advanced Manufacturing in Alabama
Disaster Medicine and Public Health Preparedness	Cambridge University Press	U.S. Governmental Spending for Disaster-Related Research
Sustainability	Environmental Research: Infrastructure and Sustainability	United States Federal Contracting and Pollution Prevention
Nonprofit Sector	Syracuse University	What Big Data Can Tell Us About Government Awards to the Nonprofit Sector
Education	Heritage Foundation	Pandemic Education Spending
Public Health	British Medical Journal Medicine	Use of Private Management Consultants in Public Health
Political Science	Cambridge University Press	Implementing Presidential Particularism: Bureaucracy and the Distribution of Federal Grants

Journalists have used open spending data to inform the public. News stories on contract and grant spending serve a variety of purposes and foster civil society. Local news stories highlight federal dollars and jobs spent in communities. News stories can spur debate about government spending priorities.

Table 2. Sample of News Stories Based on Open Data

Money Matters: Who Were USAID's Top Grantees in 2021? - *Devex*
Wyoming Left Out of Federal Coal Community Assistance Program - *WyoFile*
Nearly a Third of All Pentagon Contracts Have Gone to 5 Major Weapon Contractors - *The Boston Globe*
Government Says Contract for Covid-19 Database Was Competitively Bid - *The New York Times*

Architecture of Federal Award Reporting Systems

The data for USASpending.gov come from three primary sources. The Federal Procurement Data System–Next Generation (FPDS–NG) provides the federal prime contracts data; the Federal Assistance Awards Data System (FAADS) provides the federal assistance data, which is submitted to the Data Submission and Validation Tool (DSVT) hosted by the General Services Administration (GSA); and the FFATA Subaward Reporting System (FSRS) provides first-tier subaward data. Individual agencies report data on prime contracts and federal assistance. However, the prime grantees are responsible for reporting their subgrants data to FSRS. In addition to the primary sources of data for USASpending.gov, the website utilizes specific data sets from the CFDA and vendor/grantee registration information from the System for Award Management (SAM).

It is no longer difficult to release millions of federal award records in machine-readable formats—the technical constraints are limited. However, such advances in technology pose increased risk of mixture of good and bad quality data leading to unintended consequences. In such cases, agencies may feel that they have fulfilled their obligation of providing access to the data. In reality, an incomplete and inconsistent data set provides little added value. In fact, it may even deter third parties from expending their resources and energy on data that are incorrect, depriving the public of valuable insights.

Data Quality Challenges and Audits

Agencies use a variety of different contract writing systems and financial systems. These systems capture contract numbers, dates, dollar amounts, and other information in neatly organized databases.

Achieving high-quality, reliable data can be challenging for agencies. Ten years ago, it was estimated that 66% of data on USASpending.gov were inaccurate (Sheridan, 2011).

Agency inspector generals perform an annual validation and verification audits of procurement data (Council of Inspectors General on Integrity and Efficiency, 2020). Audits examine whether procurement and nonprocurement data element reporting is current, accurate, and timely. Internal agency data is compared against USASpending.gov reporting to verify complete and timely reporting. Accuracy is checked by comparing with supporting documentation such as official contract files (FAR 4.8, 2021). For example, the Treasury Inspector General for Tax Administration’s (TIGTA’s) Fiscal Year (FY) 2020 DATA Act Audit identified IRS procurement data elements with a relatively high error rate (Treasury Inspector General for Tax Administration [TIGTA], 2021). Table 3 provides a TIGTA Analysis of IRS DATA Act procurement and grant statistical sample transactions.

Table 3. Comparison of FY2020 and FY2019 Statistical Sample Testing Results of Reported Data Elements With Error Rates Over 20% (TIGTA, 2021)

Data Element Name	***FY 2020***	***FY 2019***	***Change***
Primary Place of Performance Address	**44%**	52%	- 8%
Potential Total Value of Award	**29%**	35%	- 6%
Primary Place of Performance Congressional District	**26%**	21%	5%
Action Date	**25%**	28%	- 3%
Legal Entity Address	**23%**	19%	4%
Current Total Value of Award	**23%**	35%	- 12%
Period of Performance Current End Date	**22%**	24%	- 2%
Period of Performance Potential End Date	**21%**	28%	- 7%
Ultimate Parent Legal Entity Name	**21%**	52%	- 31%
Ultimate Parent Unique Identifier	**21%**	23%	- 2%

Automating DATA Act Validations

Given the scale of federal financial transactions—maintaining reliable, high-quality financial data can be challenging. The use of emerging technologies such as RPA and natural language processing can reduce manual work for agency employees and improve the consistency of financial data. RPA bots emulate humans in performing computer tasks (e.g., bot clicks with mouse and types information in systems). Further intelligent automation adds complex, artificial intelligence (AI) reasoning capabilities (e.g., locate specific information in varying locations within contract documents). These technologies are key to success on financial audits and maintaining public confidence in the reliability of contract and noncontract financial information.

The FPDS has a number of built-in data validation rules. When contract specialist users input invalid information that conflicts with an observable rule, the contract action record will be held in a draft status and prevented from publishing. Additionally, when mandatory data elements are left blank, the system will also prevent publication of contract action requires with missing data.

Table 4. Data Validation Rules Pulled from the FPDS User Interface on March 7, 2022

Sample Business Rule Validations	**Sample Missing Data Validations**
1. If the transaction is an initial award, then positions 7 and 8 of the PIID must be equal to the fiscal year of the date signed. 2. If the "Date Signed" on the action is on or later than February 3, 2017, only the values "Consolidated Requirements," "Consolidated Requirements with Written Determination," "Consolidated Requirements Under FAR 7.107-1(b) Exception," "Not Consolidated" can be selected for the Data Element "Consolidated Contract." 3. "Place of Manufacture" can only be "Mfg in U.S.," "Mfg outside U.S. - Use outside the U.S.," "Mfg outside U.S. - Resale," "Mfg outside U.S. - Trade Agreements," "Mfg outside U.S. - Commercial information technology," "Mfg outside U.S - Public interest determination," "Mfg outside U.S. - Domestic non-availability," "Mfg outside U.S. - Unreasonable cost," or "Mfg outside U.S. - Qualifying country (DOD only)" when the "PSC Code" is "1000-9999," except the values: 5510, 87**, 88**, 89**, 9410, 9430, 9440, 9610, 9620 or 9630.	1. Mandatory element: "nationalInterestActionCode" is missing for the award. 2. Mandatory element: "Emergency Acquisition" is missing for the award. 3. Mandatory element: "signedDate" is missing for the award.

A growing marketplace of vendors and technology solutions is helping to improve federal contract and financial data quality.[63] For example, FedDataCheck (www.feddatacheck.com) is an automation solution that runs additional validation checks and emails contract specialists requesting correction of data that appear to be erroneous. Having an automated data validation tool assists agencies in conducting FPDS validation and verification activities.

[63] References to brand names and vendors are provided to assist agencies in finding products suitable for meeting agency needs. No endorsement is implied. This paper describes the salient functional performance characteristics of data validation software products to support acquisition planning in accordance with FAR 10.001(a)(3)(ii) and FAR 11.104.[64] Such as maximizing output quantity, subject to input constraints.

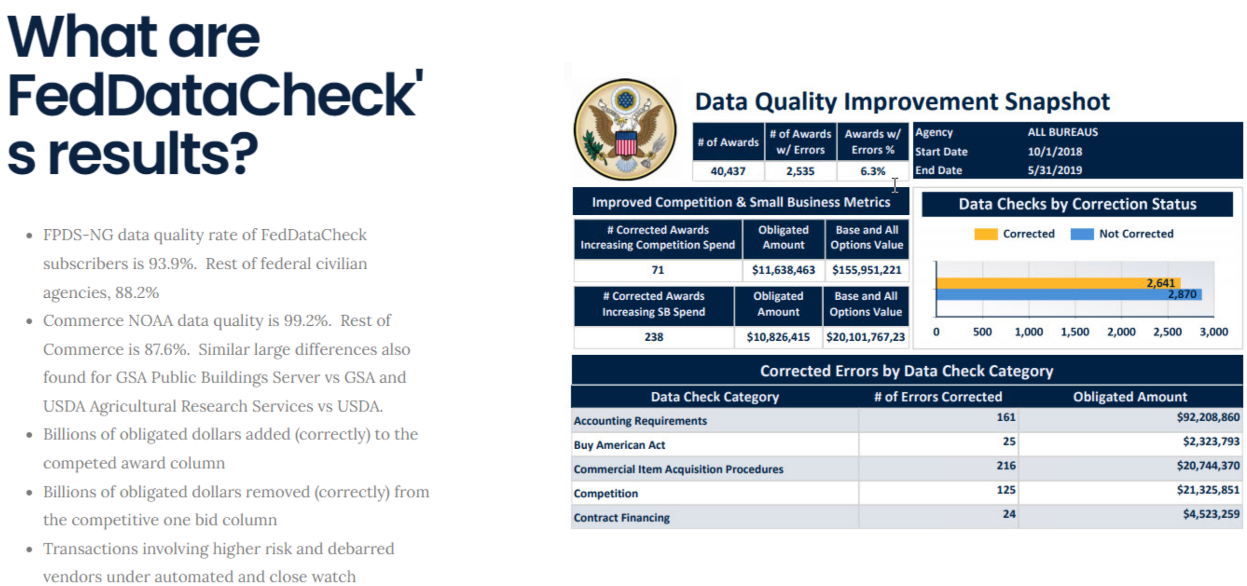
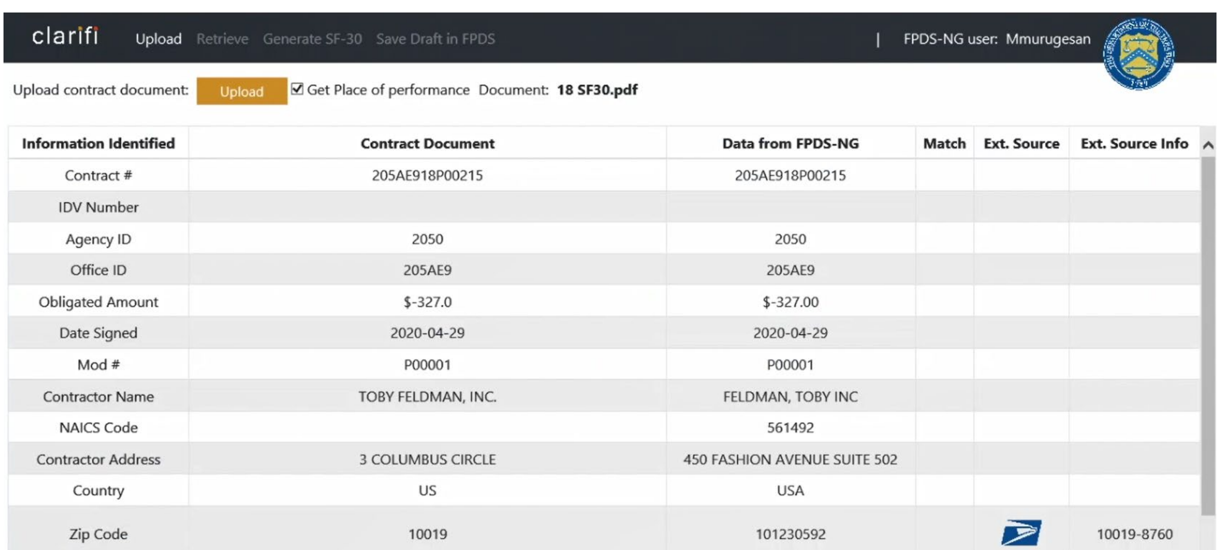


Figure 1. What Are FedDataCheck's Results?

The IRS held a competition seeking solutions to further improve data quality. Goals of the acquisition were to "achieve incremental improvement in IRS data" and "limit the amount of manual work required by government personnel" (G2Xchange FedCiv, 2019). Five vendors were selected to participate in an IRS Data Act Pilot. Significant progress had been made in improving the quality of FPDS data, but audits indicated that database entries did not always match information in signed contract award and modification documents.

Intelligent automation can locate specific pieces of information in contract documents. Data elements extracted from contract documents, such as dollar amounts, dates, and addresses, can then be compared with corresponding database records for consistency. The below screenshot shows the DATA Act bot in action. A PDF format contract modification was downloaded from a contract writing information system by the bot. DATA Act information was extracted from the contract modification document and compared for consistency with the corresponding FPDS contract action report.



Information Identified	Contract Document	Data from FPDS-NG	Match	Ext. Source	Ext. Source Info
Contract #	205AE918P00215	205AE918P00215			
IDV Number					
Agency ID	2050	2050			
Office ID	205AE9	205AE9			
Obligated Amount	$-327.0	$-327.00			
Date Signed	2020-04-29	2020-04-29			
Mod #	P00001	P00001			
Contractor Name	TOBY FELDMAN, INC.	FELDMAN, TOBY INC			
NAICS Code		561492			
Contractor Address	3 COLUMBUS CIRCLE	450 FASHION AVENUE SUITE 502			
Country	US	USA			
Zip Code	10019	101230592			10019-8760

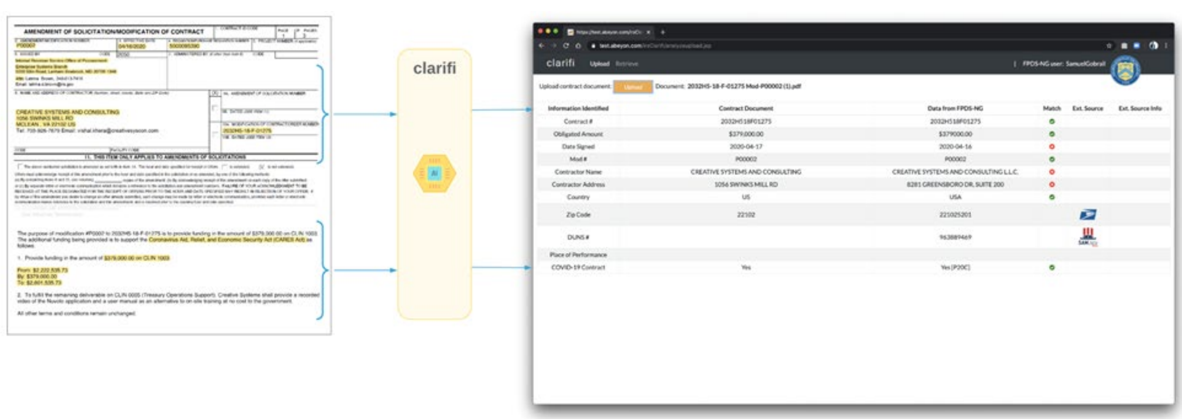
The DATA Act bot utilizes AI as well as RPA. These technologies combined allow for the automation of the process in four major stages. The first stage is to properly identify the relevant contract document that will contain the valid source of data elements, amongst nonrelevant procurement documents within a procurement folder management system. The second stage is to dive into the identified document and identify structured and unstructured data elements that are aligned to the FPDS–NG record. The solution then compares the extracted data to the data present in FPDS–NG and identifies all discrepancies in a report. Finally, through the report interface, the human administrator can choose to have the bot resolve all or specific discrepancies identified within FPDS–NG.

DATA Act Bot Stage 1: Extraction of the Correct Documents

The DATA Act bot success relies heavily on the identification of the correct document, replicating the ability of a human administrator to sieve through many documents and utilize historical knowledge to select the document that has the highest likelihood of containing the information pursued, based on training of the AI models implemented. Within the IRS, for example, procurement documents are filed electronically in accordance with a specific contract filing checklist, which varies depending on the stage of the procurement and type of procurement that the documents are associated with. The automated solution must determine the appropriate location for each document according to the filing path of each checklist and upload the document into the solution interface for Stage 2 utilizing RPA.

DATA Act Bot Stage 2: Data Elements Extraction

Stage 2 of the solution involves the extraction of data elements from the documents selected in Stage 1, across structured and unstructured fields. The solution is designed to extract 32 data elements ranging from **Date Signed**, an important field capturing the data of the procurement action's execution, to more complex elements such as **Period of Performance**. The complexity of each data element depends heavily on two factors: (1) data element availability and (2) the predictability of the location of the data element, as in, the frequency of the data element appearing in the same location(s) of the document every time. Data element availability refers to presence of this data element across a variety of different documents. A good example of this is **Contract Number** or **Vendor Name**, where the data can be found in documents ranging from official procurement forms (SF30s, SF1449, etc.) to Statement of Work documents, whereas data elements such as **NAICS Code** and **Obligated Amount** are limited to specific documents. AI models for complex data elements would have to properly identify the proper document where they would reside as well as the exact location of such data element within the document, considering human variability of data input.



Accuracy Assessment

As previously mentioned, to automate the validation of FPDS–NG records, the automation must be able to detect data elements from source data with a high degree of accuracy, comparable to that of a human administrator. Consider Acceptable Quality Levels (AQLs) typically included in contract Quality Assurance Surveillance Plans. An iterative, incremental process was used to improve model performance on each FPDS data element. Data validation models are given additional data, or changes are made to machine learning algorithm hyperparameters until an AQL is reached. F1 scores are calculated for each data variable (i.e., data field in FPDS) to determine accuracy and reliability of the model.

DATA Act Bot Stage 3: Reporting Discrepancies

Once the data elements are properly identified from the source procurement document, they are extracted and compared to the FPDS–NG data elements to identify discrepancies. The results are then downloaded and presented to a human administrator with discrepancies highlighted for validation. The administrator can select within the report which discrepancy should be corrected by the automation. This is done through a simple prompt in the report.

DATA Act Bot Stage 4: Resolving FPDS–NG Discrepancies

The marked discrepancies triggered by the administrator would prompt the automation to log onto FPDS–NG and resolve the discrepancies by modifying the data element in FPDS–NG to the data found in the procurement documents. The modified record could be saved in either draft mode or final mode depending on user and system input.

DATA Act Bot: Results

The DATA Act bot has brought demonstratable improvement to agency performance on DATA Act audits. The bot automates the tedious work of verifying the consistency of contract dates, dollar amounts, addresses, and other information. The TIGTA's FY2020 DATA Act Audit states that the "IRS received a score of 97.7 based on our sample and therefore has an overall quality rating of 'Excellent'" (TIGTA, 2021). TIGTA recommended that the IRS continue automated quality review efforts.

Conclusion

The United States has a robust program for spending transparency that has already brought significant benefits. That said, opportunities for improvement exist. Greater use of automated data validation with business rules, natural language processing, and machine learning will increase data quality and transparency. A 2021 Government Accountability Office (GAO) report found that USASpending.gov fosters greater transparency, informs decision-making, and helps identify fraud (Government Accountability Office, 2021). Further, GAO noted that the website employs user-friendly, human-centered design. GAO recommended increasing training for targeted user groups to obtain more benefits.

References

Council of Inspectors General on Integrity and Efficiency. (2020, December 4). *Inspectors General guide to compliance under the DATA Act*. https://oig.treasury.gov/sites/oig/files/2021-01/OIG-CA-21-008.pdf

Digital Accountability and Transparency Act, Pub. L. No. 113-101 (2014). https://www.congress.gov/113/plaws/publ101/PLAW-113publ101.pdf

FAR 4.8, Government Contract Files (2021). https://www.acquisition.gov/far/part-4#FAR_Subpart_4_8

Government Accountability Office. (2021, December). *Federal spending transparency: Opportunities exist for Treasury to further improve USAspending.gov's use and usefulness* (GAO-22-104127). https://www.gao.gov/products/gao-22-104127

General Accounting Office. (1980, April). *The Federal Procurement Data System: Making it work better*. HeinOnline.

Government Contractor. (2006). President signs act creating web site of contract and grant recipients. *Government Contractor*, *48*(345).

Federal Housekeeping Statute, 5 U.S.C. § 301 (1789).

Huijboom, N. (2011). Open data: An international comparison of strategies. *European Journal of ePractice*. https://research.vu.nl/en/publications/open-data-an-international-comparison-of-strategies

Maurushat, A., Trzaskowski, J., & Al-Alosi, H. (2017). Open data: Turning data into information, and information into insights that allow for evidence-based policy. *Newcastle Law Review*, *12*, 104–147.

G2Xchange FedCiv. (2019, October 1). *IRS awards Data Act Improvement Pilot to five companies*. MileMarker10. https://etc.g2xchange.com/statics/irs-awards-data-act-improvement-pilot-to-five-companies

Rizvi, Z. (2016). Dubai data law. *Court Uncourt*, *3*(3), 2–4.

Federal Funding Accountability and Transparency Act of 2006, Pub. L. No. 109-282 (2006). https://www.congress.gov/congressional-report/109th-congress/senate-report/329/1

Sheridan, B. (2011). *DATA Act: The federal case for XBLR*. Maryland Association of Certified Public Accountants.

Treasury Inspector General for Tax Administration. (2021, October). *Fiscal Year 2020 Digital Accountability and Transparency Act reporting compliance* (Report No. 2022-10-003). https://www.treasury.gov/tigta/auditreports/2022reports/202210003fr.pdf

Yu, H., & Robinson, D. G. (2012). The new ambiguity of "open government." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2012489

Telling Time: Getting Relevant Data for Acquisition Schedule Estimating Relationships

Dr. Charles Pickar is a member of the NPS faculty, where he teaches project management, defense acquisition, and systems engineering. Before joining NPS, he led the Applied Systems Engineering Program Area at the Johns Hopkins University Applied Physics Laboratory. He is a retired Army officer with extensive experience in the U.S. defense industry—including Director and V.P. levels at Lockheed Martin, Northrop Grumman, and SAIC. He is the current Chair of the Systems Education Technical Committee of the IEEE Systems Council. His research and published work focus on applying systems engineer and system dynamics analytical approaches to defense acquisition problems. [ckpickar@nps.edu]

Brig Gen Raymond Franck, USAF (Ret.) retired from the faculty of Graduate School of Business & Public Policy, NPS, in 2012. He retired from the Air Force in 2000 in the grade of Brigadier General. His active-duty career included operational, staff, and leadership assignments. His final assignment was as Head of the Department of Economics and Geography, United States Air Force Academy. His published work includes journal articles and research reports on military innovation and defense acquisition management. [cfranck215@aol.com]

Abstract

This paper is part of a research agenda outlined in Franck et al. (2016) directed toward improving the realism of defense acquisition schedules. Defense acquisition schedules have long been a difficult problem. In this particular effort, we consider primarily the case of the 737MAX, which has been a fortuitous example of the risks of scheduling-by-fiat. We analyze the 737MAX misadventure using systems dynamics and root cause analysis methods.

A fundamental question for defense acquisition schedule estimating is the extent to which schedule drivers vary (or don't) across various defense acquisition programs. If the programs are, in fact, idiosyncratic in nature, then we have prospects of explaining observed schedules (with program-specific explanatory variables). However, to the extent that common themes drive schedules across whole classes of programs, we have better prospects of predicting expected schedule length. This paper aims to (a) present a useful perspective of this question and (b) offer suggestions for the way forward.

Keywords: Acquisition Schedules, Data Science

Why Estimating Acquisition Schedules Is Difficult

Schedule is the least understood of the three critical outcomes in weapons system development (cost, schedule, and performance) by both researchers and practitioners. As much art as science, scheduling is an aspect of the decision-making necessary to develop and deliver combat capability. The science is driven by the necessity to accurately capture the elements of the schedule to provide an accurate starting as well as measurement to the program.

Acquisition schedules have long been identified as a troublesome issue (e.g., Peck & Scherer, 1962, ch. 16). And the art of estimating schedules (or explaining schedules) has received decades of attention since. One approach to this problem has been schedule estimating relationships (SERs), which posit an orderly relationship between actual schedule and observable (hopefully quantifiable) factors relevant to any given program (Franck et al., 2016, pp. 99–100). Franck and colleagues also offered a preliminary list of explanatory variables for an SER.

Schedule estimates (ex-ante) and schedule analysis (ex-post) are easier said than done. They involve both art and science. The "science" part includes a systematic study of the relevant data, often distilled into quantitative relationships. The "art" aspect arises, inter alia, from the inherent complications in any endeavor with a schedule. And the discussion that follows provides insights into complications attendant to any schedule estimation.

As Pickar and Franck (2019) pointed out, there are hazards to schedule estimates without a reasonable grounding in past experience. We think a promising approach toward that end is SERs, well informed by experience. If properly developed and applied, this tool can significantly improve life for acquisition professionals.

It might also improve the lot of those who study acquisition matters. For example, Trudelle et al. (2017) studied how likely defense acquisition programs were to stay within cost and schedule bounds. A vexing issue in this effort was the dubious reasonableness of the original schedule estimates from which they measured overruns. Discussing the issue of major defense acquisition programs having a greater likelihood of leaving estimated bounds, they encountered the issue of whether the initial schedule (and cost) estimates were a matter of reasonably confident expectation: Addressing the common practice of "optimistic" initial estimates, the authors note, "We have seen too many . . . schedule times (that) are simply unrealistic" (p. 611).

Some Near-Universal Schedule Estimating Complications

The "Incidentals": A Road Trip Schedule Estimating Problem

"People always fail to plan for the incidentals."

Time to complete (schedule) can be affected by factors external to the program. Consider a simple example. Suppose a firm (A) specializes in hauling small, high-value cargoes over relatively long distances by road. Suppose also "A" is bidding for a contract to pick up cargo at San Antonio International Airport and deliver it to Chicago O'Hare on a date six months in the future, with time en route a significant factor in the contract award. A confirmed optimist would note that the great circle distance between those points is 1,040 (statute) miles—which (if practical) implies about 15 hours of driving time (averaging 70 mph). And a confirmed optimist might well propose that as an estimate. A more realistic estimate is to pick the fastest route with actual roads—which turns out to be almost entirely interstate highways. This works out to 1,250 miles—about 18 hours. (See Figure 1.)

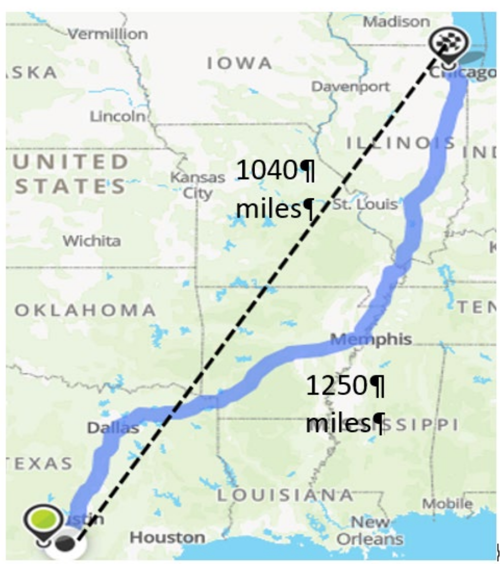


Figure 1. Road Trip Schedule

The second estimate is more credible, but there's a tendency to underestimate the role of incidentals. These include stops for gas and other things, plus delays due to rush-hour traffic. Planning factors for this set of variables are relatively easy to formulate. However, the possibilities of mechanical breakdown or mishap are more difficult.

Also, there are external factors that may arise. A delay in package arrival in San Antonio could change the rush-hour delays. Road construction would also be a factor (and probably not easily predicted six months ahead). One way to improve the estimate might be reliance on data from previous trips of this nature.

As Riposo et al. (2014, p. 41) point out, schedule drivers include factors outside the program itself. These include funding stability (or not). Other external factors include the following:

- acquisition policy regime (McNicol & Wu, 2014),
- funding "climate" (McNicol, 2015, 2020),
- external shocks, such as significant funding changes, new requirements (GAO, 2010), bid protests with associated litigation (Amara & Franck, 2021), and
- "acts of God" (such as hurricanes; Werner, 2019).

Managed Processes and Outside Observers

Processes whose outcomes are influenced by management actions are more complicated than those determined by nature or a simple optimization process.[64] Such complications can arise when program management must make trade-offs among multiple outcomes, such as cost, performance, and schedule.[65]

Large projects entail significant management effort—to avoid inefficiencies and make appropriate balancing of multiple goals. However, even straightforward projects lead to trade-offs and complications—particularly for prognosticators.

Consider a very simple project consisting of two tasks, M and N. This is summarized in Figure 2. The project tasks are accomplished sequentially by two teams (Teams M and N, respectively).

Model variables determined by nature are the following:

- Time to complete Task M (TCM) is 1 and time to complete Task N (TCN) is 2, each with probability 0.5, determined independently (known unknowns).
- TCM (TCN) is assumed known for one period after Task M (Task N) is started.

[64] Such as maximizing output quantity, subject to input constraints.
[65] And as the *Manual for the Operation of the Joint Capabilities Integration and Development System* (CJCS, 2018) and related directives make plain, program managers are expected to do just that.

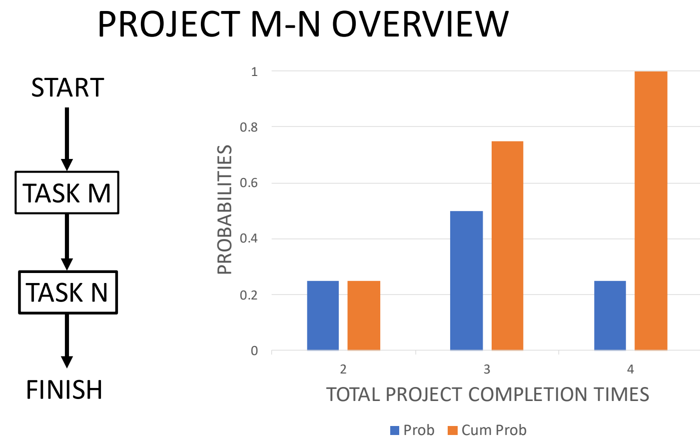


Figure 2. Project M–N

Note: Probabilities of project completion times assume no delays. (Incurring risk of delay is a management decision.)

The manager's decision variables are the following:

- TPN: when Team N is assembled and ready for Task N (TPN = 1 or 2);
- TM (TN): time spent with Task M (N) = 1 or 2.

Outcomes depend on task completion times (TCM, TCN), times spent on each task (TM, TN) and TPN, with

- Schedule = TM + TN + Delay[66]
- Cost = TM + TN + Wait[67]
- Wait = 1, if TM = 2 and TPN = 1.
- Delay = 1, if TM = 1 and TPN = 2.
- Performance ≈ [(TM/TCM)^0.5 + (TN/TCN)^0.5] * 50.[68]

Management has three decision variables: TM, TN, and TPN. Setting TPN = 1 assumes a risk of cost increase if TM = 2, causing Team N to spend one period idle (but paid). If TM < TCM or TN < TCN, then the management sacrifices performance in favor of schedule and cost.

---

[66] Schedule can vary between 2 (TM = TN = TPN = 1) and 4 (TM = TN = 2). If TPN = 2, and TCM = 1, then one period passes with no work done, a wait which adds one period to completion time (scheduled).
[67] Cost can vary between 2 (TCM = TCN = 1) and 5 (TM = TN = 2; and TPN = 1). If TM = 2 and TPN = 1, then Team N is waiting for Task M to complete. This adds one unit of cost.
[68] Performance of the developed product depends on time allocated to each task (M,N) versus time to complete the phase.

Project Management Scenario

As we have already noted, project management is expected to care about performance, cost, and time to complete (schedule). The management time line looks like Table 1.

Table 1. Project Management Decision Sequence When Managing for Performance

Time	Information	Decision Variable(s)	Outcomes Emerging
0	Initial set (above)	TPN	
1	TCM (1 or 2)	TM (1 or 2)	Performance partially defined. Cost, schedule choices narrowed.
TM + Delay + 1	TCN (1 or 2)	TN (1 or 2)	All outcomes determined (performance, schedule, and cost).

Management starts with the information summarized above and must decide at Time 0 when to have the resources for Task N (Team N) in place. If TM = 1 and TPN = 2, then the work stops one period before Task N begins—with attendant schedule implications (delay of one period). If TM = 2 and TPN = 1, then Team N must wait until Task M is complete—with attendant cost implications (Team N in place but idle for one period).

In the rest of this section, we examine the managerial “trade space” if the M–N Project is managed (a) to meet a performance goal, (b) to stay within cost constraints (a budget), or (c) to complete by a specified time (schedule).

In the discussion below, we assume the project is managed for performance, cost, or schedule. That is, performance (cost, schedule) is fixed with cost and schedule (performance and schedule, performance and cost) varied. This leads to an efficient set for the other outcomes. For example, if the project is managed for performance, there is an efficient set of cost–schedule outcome pairs. This can be plotted as a curve, shifting the efficient-set curve as performance requirement changes.

Management for Performance

As noted, the performance achieved can vary from 70 to 100. If TCM = TCN = 2, and TM = TN = 1, then Performance = 70. If TM = TCM and TN = TCN, then performance is 100.

Management strategy depends on the performance goal specified. If Performance must be 100, then, of course, TM = TCM, and TN = TCN. Cost and schedule are then determined if the observer also knows TPN. If project management strongly emphasizes cost (schedule) over schedule (cost), then TPN = 2 (1).

If the performance requirement is less than 100, the program manager (PM) should set TM = 1—with TN = TCN to reach (or exceed) the performance goal even if TCM = 2. Knowing that TM = 1 is a given, then TPN = 1 is preferred since that’s when Task N will start (regardless of TCM). Management can choose TN to reach the goal (and benefits if TCN = 1).[69]

---

[69] If performance requirement is 70, and either TCM or TCN = 1, then product performance must exceed that requirement.

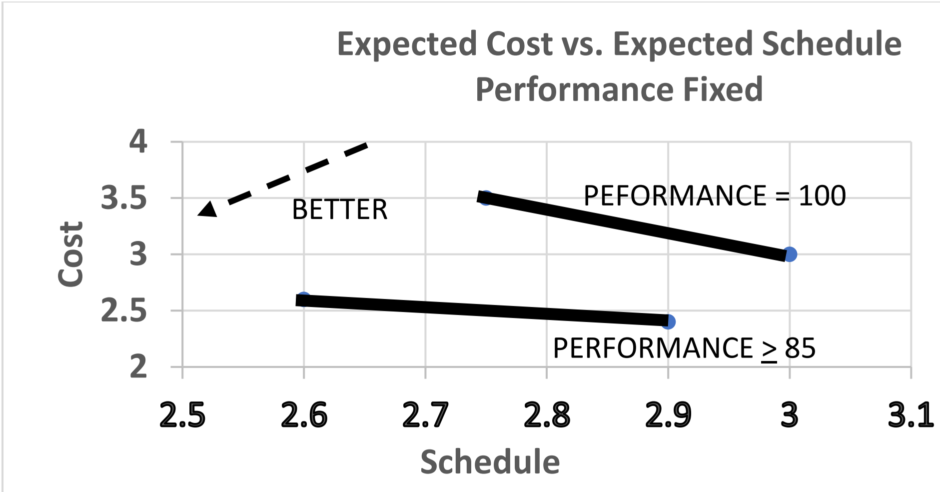


Figure 3. Managing for Performance

Note: Although extending the schedule by one time period automatically increases cost by 1, there's still a cost–schedule trade-off in this example.

Managing For Cost (Fixed Budget)

The trade-off between performance and schedule, with various cost limits, is shown in Figure 3. Clearly, striving for better performance increases both cost and schedule time—or both. Cost can vary from 2 to 5. If TM = TN = 2 and TPN = 1, then Cost = 5. If TM = TN = 1, then Cost = 2.

In this case, since the project budget is fixed, then management must consider trade-offs between performance and schedule. Cost can vary from 2 (TCM = TCN = 1) to 5 (TCM = TCN = 2, and TPN = 1, resulting from Team N waiting one period).

<u>If Cost (Budget) is fixed at 2</u>, then the project must pursue that low-confidence, success-oriented strategy—with TM = TN = 1, with (of course) TPN = 1. If this does pan out, then the project will be highly successful with low cost (2), quick completion (2), and excellent performance (100).[70]

If either (or both) task completion times are 2 (Prob = 0.75), then there is a performance penalty, summarized in Table 2. Since budget dictates schedule, there is no schedule or delay risk accepted in the PM's strategy. (The one path to success entails TPN = 1, with no schedule or cost penalty incurred.)

Table 2. Performance Outcomes With Best Management Strategy and Budget = 2.

	TCN	
TCM	1	2
1	Performance = 100	85
2	85	70

[70] Enthusiasts and optimists tend to tout this as the mostly likely result.

If Cost = 3, then decision space increases with the ability to deal with longer task completion times. An excellent original plan is TM = TCM, with TPN = 2 (to save on expected cost).

If TCM = 1 = TM, then there is a delay (waiting for Team N to get in place). Following that event, there is a performance–schedule trade-off if TCN = 2.

If TCM = 2 = TM, there is likewise a performance–schedule trade-off. At TM = 1, then the TN (decision variable) versus TCN (determined by chance) determines performance, schedule, and cost outcomes. For example, if TCN = 2, then TN = 1 saves time, while TN = 2 increases performance.

If performance is considered much more important than schedule, then expected performance, schedule, and cost are 94, 3.4, and 2.6, respectively. If schedule is considered much more important than performance, then the outcomes are 89, 3, and 2.25.

If Cost = 4, then management decision space increases yet again, with the added option of TN = 2, even if TM = TCM = 2. Once again, TPN = 2 to save money.

The increased budget buys a better trade-off between expected performance and expected schedule. If performance is much more critical that schedule, then expected performance, schedule, and cost are 96, 3.5, and 2.8, respectively. If schedule is more important than performance, then the expected outcomes are 89, 3, and 2.25.

If Budget = 5, then the project is figuratively awash in cash, and management can formulate a can't-miss strategy of TM = TCM, and TN = TCN, with TPN = 1—with the project cost of 2 to 5 (depending on TCN). The schedule–performance trade-off is painless in terms of meeting guidance.

Maximum achievable performance is (of course) 100 with expected completion after three periods. Performance of 85 is possible with TM = 1, TPN = 1, and TN = TCN—with an expected completion time of 2.5. Performance of 70 can be attained with TM = TN = TPN = 1, with the expected schedule of 2. However, performance exceeds 70 with a probability of 0.75 (with an expected performance of 84), even if TM = TN = 1.

The trade-off curve for performance and schedule is shown in Figure 4. Given schedule, a higher budget increases expected performance. Given performance, a higher budget enables a shorter schedule.

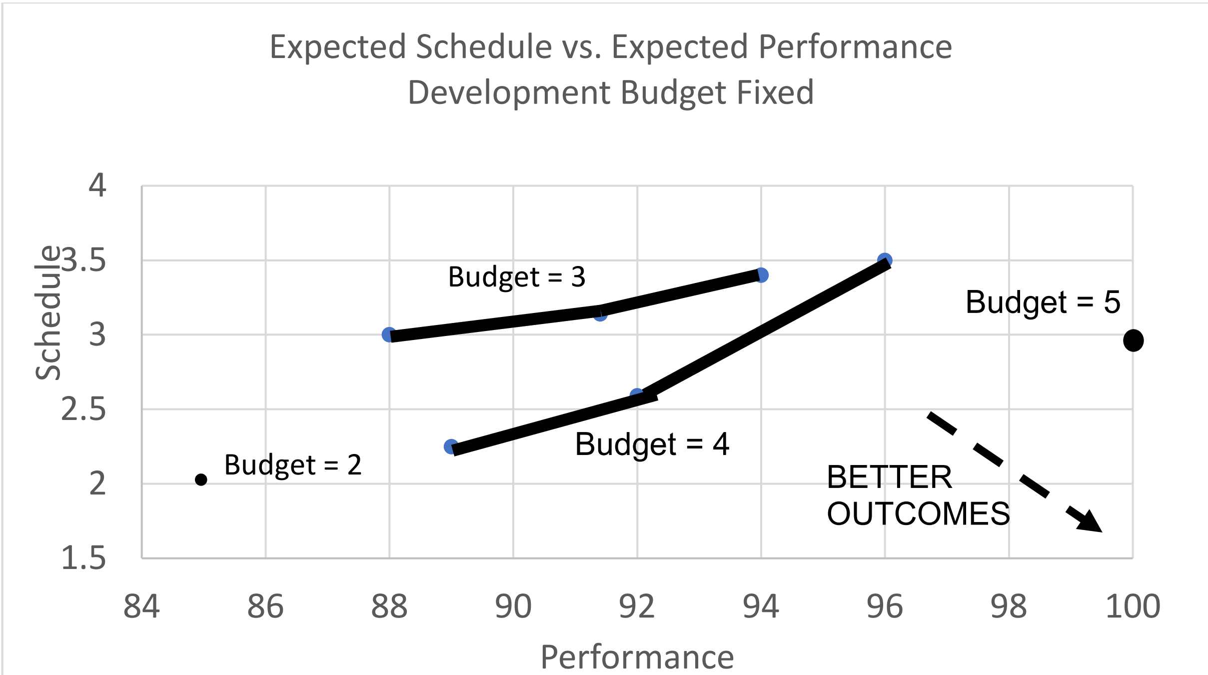


Figure 4. Managing for Development Budget (Cost)

Managing for Schedule

Schedule can vary from 2 to 4. If TCM = TCN = 1 and TPN = 1, then Schedule = 2. If TCM = TCN = 2, TM = TM = 2, then Schedule = 4 (also if TM = 1, TPN = 2, and TN = 2).

If Schedule = 2, there is only one path to success: TCM = TCN = TM = 1 and TPN = 1 (as discussed above for cost restricted to 2).

If Schedule = 3, TPN = 1—to save time at risk of increased cost (due to Team N possibly assembled and waiting).

If Schedule = 4, TM = TCM, TN = TCN, and TPN = 2 (to remove a risk to cost).

Schedule versus performance trade-off curves are shown in Figure 5.

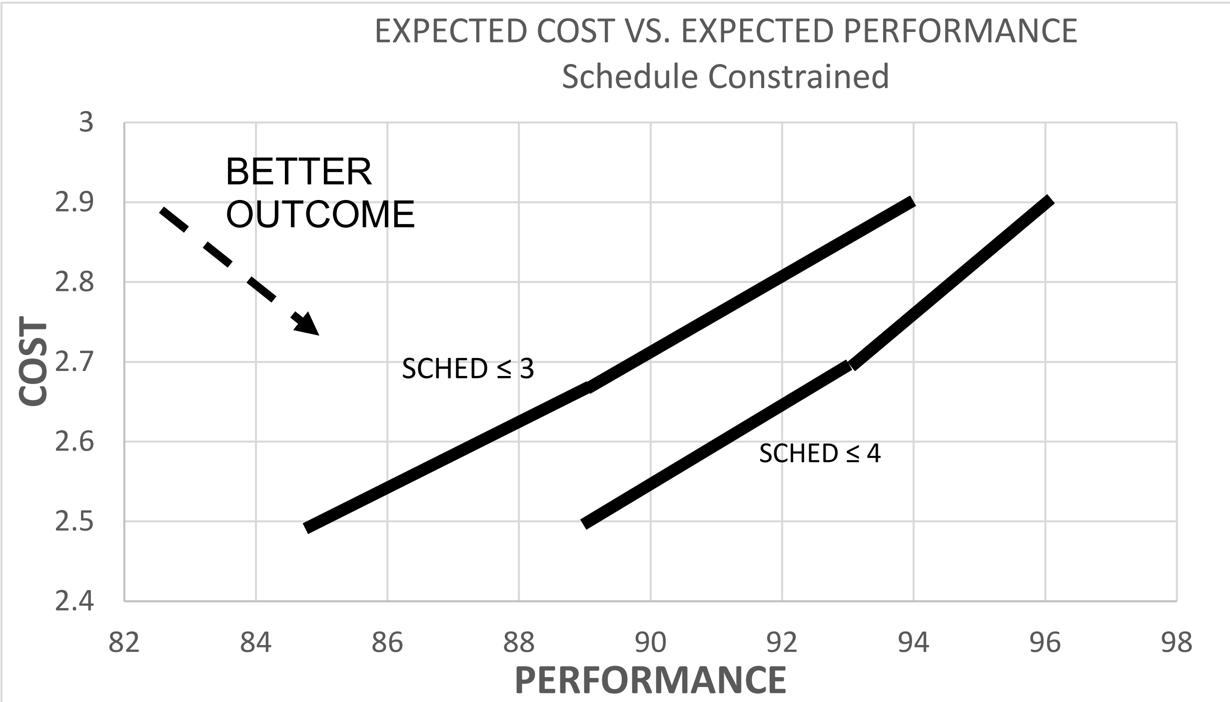


Figure 5. Managing for Schedule

Even a straightforward managed process can pose complications for analysts and forecasters. Any one of the three Joint Capabilities Integration and Development System (JCIDS) macro-outcomes can be changed by management decisions. Thus, estimating schedule means, among other things, predicting PM decisions. Even knowing in advance TCM and TCN is not enough to predict schedule (unless, of course, program management also knows them in advance).

Why this is so: Referring back to Figure 1, suppose performance must be 100. Then management seeking to lessen costs would minimize cost subject to the performance constraint to achieve an expected schedule of 3 and expected cost of 3.

If, however, program management emphasizes schedule, then the expected schedule is minimized subject to the performance constraint (100). In this case, the expected cost is 3.5, and the predicted schedule is 2.75. Note that management decisions determine schedule, as do the unknowns (TCM, TCN). Therefore, an ex ante estimate of program schedule (even with detailed prior knowledge) encounters opaque factors to outside observers using current methods for formulating estimating relationships. In short, actual schedules also depend on decisions the PM makes.

Complications of managed processes appear in the defense acquisition literature. Examples follow.

- The hypothesized relationships can be complicated. For example, Light et al. (2017) included "planned concurrency" in one of their regression models. The resulting coefficient was significant and strongly against expectations (p. 26). The authors offered the highly plausible hypothesis that programs with *planned* concurrency had traits such as (relative) simplicity and were inherently robust with respect to program miscalculations (such as not much associated rework; Light et al., 2017, p. 9).

- A Government Accountability Office report in 2010 cited stable funding as a significant feature of stable acquisition programs (p. 27). The likely Department of Defense (DoD) interpretation would emphasize funding perturbations (including continuing resolutions), impeding effective program execution (e.g., Shackelford, 2021). An Armed Services Committee response (e.g., Adam Smith comment in O'Hanlon, 2021) is likely that program instability causes funding instability. This seems complicated to sort out, particularly in updating schedule estimates. This also indicates that multiple agencies involved in managing a program can indeed add additional complications.
- The literature sampled includes a fair amount of attention to development activities that occur before Milestone (MS) B is offered as an explanatory variable for cycle time from MS B to later events such as low rate initial production (LRIP) or initial operational capability (IOC; Boyd & Mundt, 1995; Harmon et al., 1989). One (perhaps naïve) view is that doing more before MS B means doing less after MS B—with time from MS B to, say, MS C obviously shortened. Another view is that activities before MS B enable a more informed source selection and associated contract—which will likely shorten the time to LRIP in any case. Perhaps both assessments are valid.

What Motivates Those Doing the Schedule Estimates?

Realistic schedule estimates just don't happen spontaneously. Examples from the literature follow.

First, Light et al. (2017), among others, focused on differences (in cost and schedule) concerning the original estimates. However, the authors also noted that those original estimates are flawed in interrelated ways. In the enthusiasm that attends the launch of a new program, there is a "tendency to believe that a current project will go as well as planned "despite previous experience in similar circumstances" (Light et al., 2017, p. 2). This has been dubbed the "planning fallacy" (Kahneman, 2011, pp. 249–251; O'Neil, 2011, p. 286).

Second, source selections have incentive structures that discourage conservative (and more realistic) estimates by prospective vendors—since realistic estimates result in being less likely to get a high-stakes contract. As one participant put it in the context of the advanced medium-range air-to-air missile (AMRAAM) source selection, "There is only one program like this every 30 years, and programs last about that long, so you're driven to go after this work. . . . Sometimes it's a matter of staying in the business" (Mayer, 1993, p. 10).

Third, while there are potentially serious risks associated with winning the contract and being unable to deliver as promised, these are encountered after the "fundamental transformation" of competitive selection to something like a bilateral monopoly (one buyer and one seller) in which even a firm in serious execution difficulties has substantial bargaining leverage (Williamson, 1996, pp. 13, 60–61). It's generally better to have the contract (even with difficulties) than being on the outside and looking in.[71]

[71] The current state of the KC-X aerial tanker program is illustrative. Boeing's KC-46 won the contract, but the KC-46 still has serious operational shortfalls a decade later. Nonetheless, recent Airbus Group offers to supply their KC-45s have not gone far, at least not yet.

In short, those who know most about the proposal in question are generally (a) highly optimistic and (b) incentivized to be optimistic.[72] This strongly discourages realistic schedule estimates at the program start.

Acquisition Programs: Commonalities Versus Differences

This is a significant issue in the study of defense acquisition schedules. As noted above, if acquisition programs are inherently sui generis, then the critical schedule drivers may not emerge until the program is well underway—which leaves (ex ante) schedule estimators a challenging task. Their lot improves if there are indeed common factors.

Our discussion begins with two studies of schedule lengths for aircraft programs. First, Harmon et al. (1989) analyzed completion times for several portions of the development process for several third- and fourth-generation fighter and attack aircraft. Their paper identified 14 candidate variables (p. 138). Variables with strong explanatory power were

- program-specific parameter,
- airframe size (empty weight),
- contractor,
- prototypes (yes or no),
- supply-chain teaming, and
- production (rate and cumulative numbers; pp. 271–278).

On the other hand, Boyd and Mundt (1995) analyzed schedules for "heavy" aircraft (bombers, transports, tankers, and surveillance) over a long period (B-29 to C-17). Useful explanatory variables were

- date of engineering and manufacturing development (EMD) start,
- airframe size (maximum thrust, number of engines, and wing area),
- combat mission (yes or no), and
- prototypes (yes or no; pp. 142–144).

Interestingly, studies of similar program types undertaken relatively close in time (1989, 1995) by researchers from the same institution (the Institute for Defense Analyses) have commonalities and differences. The commonalities include aircraft size and prototyping (or not). For example, Harmon et al. (1995) estimated that having prototype aircraft can extend, or shorten, schedules in different phases of development (pp. 271–278), while Scott and Mundt's (1989) readers seem invited to conclude that prototyping extends program schedules (pp. 142–144).[73]

---

[72] This set of incentives arises in part from the DoD's tendency toward long-term, high-value, winner-takes-all source selections.
[73] Prototyping is more likely to be observed in ambitious and complicated programs and is a method used to mitigate the inherently longer schedules of these programs. (If so, then prototyping is associated with longer schedules, but not a cause of longer schedules.) Among other things, this is a manifestation of the managed-process issue.

However, a striking indication for program individuality is the explanatory value that Harmon et al. (1989) found in “aircraft specific parameters” and specific contractor for each program.

All things considered, a somewhat ambiguous picture seems to emerge regarding the question of commonalities versus differences among programs. Harmon et al. (1989), finding the explanatory value from “program specific parameters” and “contractor,” indicated the presence of characteristics peculiar to each program.

There is likewise a mixed picture from the airframe size parameters. Harmon et al. (1989) found capture airframe size with weight, while Boyd and Mundt (1995) operationalized size with engine thrust, the number of engines, and wing area.

Finally, differences in explanatory variables likewise indicate individual differences in acquisition programs. For example, Harmon et al. (1989) found supply chain characteristics and production variables useful, while they did not appear in Boyd and Mundt’s (1995) reported model.

While it’s reasonable to conclude from the examples that individual program characteristics are more important than commonalities, others find common themes. For example, the GAO (2010) undertook interesting case studies of program stability (defined as being “on track” concerning cost and schedule; p. 2). The study found the following characteristics common to the stable programs considered:[74]

- strong senior leadership support, disciplined PMs, and solid business plans that were well-executed (p. 9);
- strong PMs who shared key attributes (prior experience, communications skills, and willingness to report bad news; p. 14);
- capability needs that are addressed in achievable increments based on well-defined requirements (p. 16);
- mature technologies and production techniques (p.19); and
- funding stability (an “essential ingredient” for a successful program; p. 27).

Riposo et al. (2014) compiled an overview of factors causing schedule delays, distilled from the relevant literature (among other things). They grouped sources of delay in the literature into major categories: requirements development, generation, and management; managing technical risk; resource allocation; defense acquisition management; and “other” (Riposo et al., 2014, p. xi). Interesting findings included the following:

- Realistic cost and schedule estimates are essential in improving schedules (pp. 58–59).
- However, incentive structures can discourage realistic estimates—especially when competing for initial funding (p. 32).
- Several studies indicate good management of technical aspects (including technical risk) is likely the most crucial part of schedule improvement (p. 56). (Riposo et al. accordingly

[74] The report equated “stability” with ”success” (pp. 10–15, 27).

viewed “schedule improvement as an objective for acquisition managers” [p. 35, ch. 3]).[75]

- Factors external to the program itself can significantly influence schedules (p. 41).

Shackelford (2021) also undertook an overview of factors common to successful (or not) defense acquisition programs. Key factors cited were

- quality of communications and degree of trust (p. 4),
- requirements and funding stability (p. 9),
- sufficient production-representative test assets before MS C (p. 12),
- good management decisions (p. 16), and
- strong, experienced program management (p. 21).

In our opinion, these are not idiosyncratic statements of general themes; we find differences in emphasis rather than differences in content. Accordingly, we essay the following synthesis of these three perspectives.

Likelihood of success is increased significantly through

- solid, executable plans with realistic cost and schedule estimates;
- disciplined requirements development;
- managing technical risk by avoiding technological leaps;
- strong, disciplined program management;
- effective communication and building trust among stakeholders; and
- resource and requirements stability (GAO, 2010; Riposo et al., 2014; Shackelford, 2021).

While this list is interesting and promising, we question how to measure the degree to which these factors are present in any given program. “Resource and requirements stability” seems straightforward. However, operationalizing these forms of stability is easier said than done: Is requirements (resource) stability something that’s present or not (a binary condition), or are there varying degrees of requirements (resource) stability for different programs?

The effects of program management performance have been investigated using proxy variables such as PMs’ experience and credentials—with mixed results. Apparently, the most readily accessible indicators are insufficient for the purpose.

Even more problematic is measuring the quality of communications and degree of trust. This suggests new approaches (which we discuss below).

What About Complexity?

Study the past if you would define the future.

[75] A very interesting idea along this line is *optimal schedule length*, which depends in part on the characteristics of the individual program (Riposo et al., 2014, pp. 35, 47–48).

—Confucius

Conspicuously missing from the major themes in the discussion is the matter of complexity. In this section, we examine the complexity of project management to relate that complexity to the critical variable of schedule intervals—defined as the time from one milestone to the next. The milestones (a start or finish of a phase in the development process) are determined by the DoD directives. The intent is to identify variables that help explain schedule behavior and provide DoD project managers the ability to manage time more effectively.

We build on a study that described and developed a methodology for extracting schedule data from selected acquisition reports (SAR; Pickar, 2018). Our current aim is to code and analyze the SARs database using computer-assisted qualitative data analysis (CAQDAS) software. This effort builds on a study started in 2018 that described and developed a methodology for extracting schedule data from the SAR databases (Pickar, 2018). The approach for this year's effort includes the following:

- review past studies on SERs and weapon system development program complexity;
- identify reasons for delays in major programs; and
- perform a system development complex system classification assessment.

This analysis will (a) review the causes of schedule delays, (b) examine the concept of project complexity and relate that to schedule delays, (c) propose a methodology for measuring complexity in weapon systems development, and (d) explain how complexity assessment can assist in defining SERs. Central to any understanding of project/program schedule performance is an appreciation of past schedule performance. The delay and complexity factors discussed in this paper have occurred in all development programs.

Prior Research

Project performance (in both practice and research) is almost exclusively measured by adherence to cost and schedule estimates developed during the project planning process. Those estimates are often optimistic and almost always wrong. There are easily understandable reasons for schedule delays, but it is difficult to apply that knowledge to new programs.

Schedule Delays as Schedule Outcomes

Drezner and Smith (1990) explored the reasons for schedule delays in the case of 10 programs with MS I dates post-1970. The explanations included budget, funding, complexity, technical difficulty, and requirements stability. A list of these project delay factors is found in Table 3.

Table 3. Factors Influencing Schedules. (Drezner & Smith, 1990).

Factors Influencing Program Schedules of 10 Programs Post-1990
Competition at the prime contractor level
Concurrency, overlap in time and effort between the development and production phases of a program
Funding adequacy/stability
Existence of prototyping
Separate contracts for each phase of the program
Priority of the program to the service relative to other ongoing programs
External guidance such as Office of the Secretary of Defense or congressional direction, reviews, restrictions, and designations
Joint management with other agencies
Program complexity or interactions with agencies external to the program
Technical difficulty
Concept stability, or stability in mission, operational concepts, and doctrine
Contractor performance changes/contract changes
External events such as inflation, earthquakes, labor strikes, etc.
Major requirements stability
Program manager turnover
Rework
Design freeze

A more comprehensive examination of the reasons for delays in development was accomplished in 2018 by examining SARs from 1997 to 2017 (Pickar, 2018). A qualitative data analysis extracted the PM schedule comments and the reason and the duration of the delay. The total number of schedule records in the available SAR database was 3,969. The data used in this study are a subset of the SAR reports of 1,224 programs from 1997 to 2017. Each program potentially had between one and 20 entries (corresponding to the 20 years period and depending on when the program was initiated, whether breaches occurred requiring more frequent SAR, and whether any schedule changes were reported).

Table 4. Schedule Delay Factors, 1997–2017

Schedule Delay Factors
Administrative changes to schedule including updates to Acquisition Programs Baselines (APB), Acquisition Decision Memorandum (ADM) changes, as well as changes resulting from Nunn–McCurdy processes and program restructuring
Technical issues
Testing delays
Delay in the availability of critical capabilities/facilities (launch vehicle/testing facilities/initial operational test and evaluation [IOT&E] units)
Budget/funding delays
Delays attributed to the contractor
Delays because of rework
External events such as inflation, earthquakes, labor strikes, etc. **(force majeure)**
Delays due to contracting/contract negotiation/award

Explanations of the Delays

- Administrative changes include schedule updates because of acquisition program baseline (APB) and acquisition decision memorandum (ADM) changes and changes including program restructuring as a function of decisions driven by Nunn–McCurdy results and program restructuring.

- Schedule changes identified those changes reported because of acknowledgment of the actual date of occurrence. These changes are also the result of receiving approval documents from milestone decision authorities to change specific dates.
- Technical schedule changes are a result of specific setbacks in technological development.
- Testing delays include both the ability to meet scheduled test dates and technical issues discovered in the conduct of testing. When the testing found a technical issue, that technical issue was also counted as a technical problem.
- Explanations that produced no apparent changes in the schedule data reflect comments in the change explanation but do not produce an actual change in the schedule. Examples include cases of achievement of IOC/full operational capability (FOC) and redesignations of milestones driven by ADM decisions.
- Delay in the availability of critical capabilities/facilities results from weather delays, including satellite launches.
- Budget/funding delays are tied to specific notes on lack of budget, decrease in budget, or changes by Congress to the particular program.
- Delays attributed to the contractor result from construction and delivery delays and delays attributed to the delivery of subcontractor materials.
- Delays because of rework reflect both quality issues, where the budgeted work must be redone to make it functional, as well as the feedback/follow-on problems caused throughout the development.
- Force majeure are external events such as inflation, earthquakes, labor strikes, and so on.
- Delays due to contracting/contract negotiation stem from either problems in negotiation, delays in approvals for request for proposal (RFP) releases, modification to contracts, or delays in awarding contracts.

Understanding the challenges of estimating weapon system schedules requires examination of those factors that historically have led to increases in the schedules. While these studies identified factors that have contributed to increased time, they fail to provide a way to use that knowledge in the planning process to anticipate the necessary schedule increases. A second factor in understanding delays is the context of the delays, which is expressed as project complexity.

Schedule Estimating Relationships

In 1980, Smith and Friedman examined the concept of weapon system acquisition intervals. The study concluded that weapons systems schedules had increased development time and that Office of the Secretary of Defense (OSD) organization changes had little effect on schedule. The study also suggested various ways to decrease development time. In 1989, Harmon et al. examined schedule data to "provide methods for assessing the reasonableness of proposed acquisition schedules for tactical aircraft programs" (p. 259). Boyd and Mundt (1995) developed SERs for nontactical aircraft and introduced considerations of "factors that do not lend themselves to being measured using a continuous scale" (p. 133). These "schedule driver" factors included qualitative metrics such as funding stability and competition. Jimenez (2016) and Jimenez et al. (2016) developed a model to predict a program's schedule based on program characteristics determined before MS B.

A 2018 RAND report developed SERs and provides a good menu of steps to conduct benchmarking (Light et al., 2018). The most recent examination of SERs was by Jardine et al. (2019). This study examined missile and radar data to create SER-specific data sets.

The data sets and processes developed for SERs have helped PMs plan and manage schedules and provide a valuable foundation. For the most part, those processes use relationships focused on measured intervals of weapon system development associated with budgetary data or physical attributes of different systems. These statistically sound findings provide high-level visibility into potential schedule intervals. We believe, however, that one of the significant contributors to schedule growth is the complexity of the systems being developed in the DoD. Therefore, it seems logical that consideration of complexity is a valuable avenue to explore.

Complexity

Complexity is the principal dynamic of 21st-century weapons system development and a measure of how difficult the management of the development of a weapon system could be. Complexity in project management refers to those organizational, informational, and technical characteristics of the project and, by extension, the project management organization and the technical staff (Baccarini, 1996).

For the project manager, organizational complexity means hiring specialists—experts in a particular field—to address those demanding aspects of a complex system that require single-person focus. Specialization exercises a limiting function on the development, in that the specialists in a project organization can address only those issues in their specific area.

As a result, project management offices (PMOs) have increased in size to meet the needs of specialization—also resulting in an increase in complexity. This has entailed a corresponding decrease in the visibility over the entire project, a “can’t see the forest for the trees” analogy from the individual’s perspective. Thus, complexity has the potential of causing a decrease in efficiency in the execution of the project, which, among other things, could manifest as increased time.

Complexity directly affects management and decisions as the more complex the system, the more information is required. This leads to a more challenging management effort and the resultant choices required. The mixture of human-sociopolitical complexity found in weapons systems development offices further adds to this complexity (Atkinson, 1999; Pinto, 2000). Finally, complexity reduces the predictability of decisions made (Sargut & McGrath, 2011).

Definitions and explanations of complexity—managerial, engineering, and technological—abound (Baccarini, 1996; Sargut & McGrath, 2011; Whitty & Maylor, 2009; Williams, 2002). From the project management perspective, Baccarini (1996) identified two elements of complexity, organizational and technological. He further subdivides these functions into differentiation and interdependency. *Differentiation* refers to projects’ varied size and structure and the organizations that manage them, while *interdependency* describes the activities between these diverse elements (Baccarini, 1996).

Williams (2002) built on the Baccarini topology and defined project complexity as categories in two key areas, structural complexity and uncertainty. Figure 6 shows the Williams topology. Structural complexity results from the number of project elements—including the people, the organizations, and the technology—coupled with how these pieces interact with their interdependencies. This combination of interactions of the varied aspects is structural complexity. Structural complexity includes scale, connectivity, organizational structure, and development objectives. Size is about the magnitude of the acquisition system and its policies, bureaucracy, and hierarchy, including the private sector side of defense acquisition.

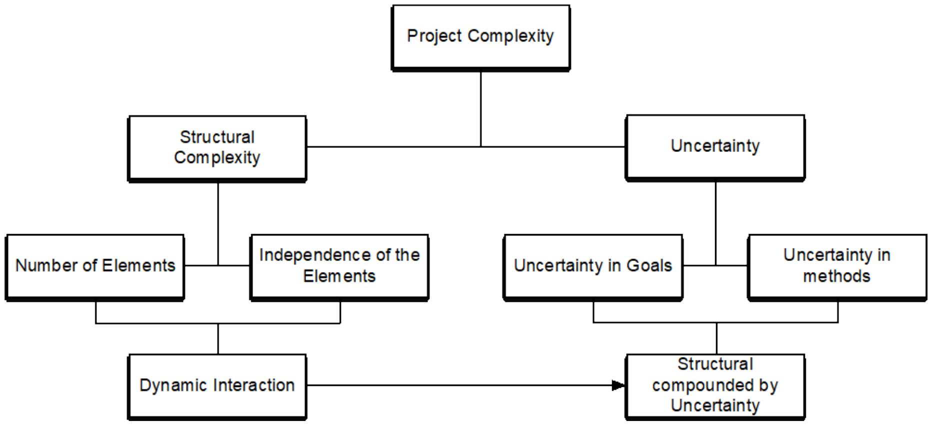


Figure 6. Project Complexity. (Williams, 2002, p. 58).

Connectivity acknowledges that the volume of staff actions between these organizations is significant and consists of both issues relating to managing ongoing development. The nature of the defense acquisition system influences the connectivity aspect of structural complexity. Since the technology development infrastructure (i.e., laboratories, research and development centers, and manufacturing) is, for the most part, privately owned, structural complexity also describes the network connectivity necessary for the system to function. Beyond the hierarchies, project organizations are major business entities directly controlling budgeting, spending, and, in most cases, the fee award to defense companies.

Project organizations are physically dispersed throughout the United States and overseas, further adding complexity. Finally, DoD project management is mirrored in the private sector by the contractor. A 2015 GAO study recognized the challenges of structural complexity in finding the reviews for some programs that include up to 56 organizations at eight levels. These structural requirements, reviews, and responding to information requests can add up to two years to the development time, significantly adding to the complexity of a development.

Uncertainty focuses on three significant areas: budget, technical complexity, and overall system objectives. Budget is a considerable concern and source of uncertainty in defense acquisition because of the year-to-year budget cycle and political considerations. Technological complexity is a fact of life in defense systems. As we develop systems, we learn more about the technologies and better plan for schedule and cost.

Sargut and McGrath (2011) identified three properties—multiplicity, interdependence, and diversity—essential to appreciate complexity. *Multiplicity* refers to the number of interacting elements or scale. This is like the Williams (2002) construct of structural complexity. *Interdependence* is the connectivity of different factors. And *diversity* is a measure of the difference in the elements (Sargut & McGrath, 2011).

Sheard and Mostashari (2009b) explained project complexity from the systems engineering perspective. The systems engineering standpoint acknowledges structural complexity but adds dynamic and sociopolitical complexity as factors influencing complex systems development (Sheard & Mostashari, 2009b). Dynamic complexity acknowledges the change-over-time of systems development. The project management system is in constant flux, whether a tactical response to a development problem or an administrative response to directives. This dynamic is a function of ongoing development's diverse and constantly changing aspects.

Sociopolitical complexity is the nexus between management, and the nonengineering human factors of policy, process, and practice of the system are most critical (Maier, 1995). Sociopolitical complexity also recognizes the politics of project management, starting with the budget process, through Congress, and back into the development organizations.

To provide an overall view and the elements of a complexity assessment tool, the complexity frameworks are summarized in Table 5. The resulting framework includes a typology of different kinds of complexity: structural, uncertainty, dynamic, sociopolitical, and overall system complexity.

Table 5. Project Management Complexity

Type	Subtype	Acquisition Management Example
Structural (Williams, 2002)	Size	Organization (number of people) Scope of work Contractor (size and number of people)
	Connectivity	Requirements organizations Industry organization Review processes (both programmatic and technical)
	Organizational	Stakeholder organizations Boundaries/different commands/different agencies Level of authority Congress
Uncertainty (Williams, 2002)	Budget	Funding
	Technical complexity	Variety of tasks Interdependencies between tasks
	Objectives	System requirements
Dynamic (Sheard & Mostashari, 2009a)	Short-term	Daily problems Personnel changeover Engineer shortage Materials failures Short requirement dynamics Rework
	Long-term	Changing budget Environment
Sociopolitical (Maier, 1995)	Human dimension	Personnel changeover Change and change management Regulations/policy changes

Measuring Program Complexity

Describing complexity is simpler than devising a means to measure it. Using the complexity breakdown above, the next step in this research is to build an assessment tool to classify selected existing weapons systems. Magee and de Weck (2004) developed a method to classify complex systems. This approach was a top-down, bottom-up review to identify and distinguish between complex systems and engineering systems. While their purpose was to differentiate complex engineering systems from traditional engineering, some elements of this approach can help classify defense systems. Similarly, Thamhain (2005) believed a tool that can determine project complexity can be valuable to the project manager as a comparative measure. Researchers in architecture and construction have also developed tools to measure complexity (Dao et al., 2016; Sinha et al., 2006).

Bosch-Rekveldt et al. (2011) developed one of the more refined studies on complexity metrics. The technical, organizational, environmental (TOE) framework consists of 40 elements shown in Table 6.

Table 6. Technical, Organizational, and Environmental Framework

Bosch-Rekveldt et al. (2011) Complexity Metrics	
Number of goals	Size of project team
Goal alignment	Size of site area
Clarity of goals	Number of locations
Scope largeness	Resource and skills availability
Uncertainties in scope	Experience with parties involved
Quality requirements	HSSE awareness
Number of tasks	Interfaces between different disciplines
Variety of tasks	Number of financial resources
Dependencies between tasks	Contract types
Uncertainty in methods	Number of different nationalities
Interrelations between technical processes	Number of different languages
Conflicting norms and standards	Cooperation JV partner
Newness of technology (worldwide)	Overlapping office hours
Experience with technology	Trust in project team
Technical risks	Trust in contractor
Project duration	Organizational risks
Compatibility of different project	Number of stakeholders
Political influence	Variety of stakeholders' perspectives
Size in engineering hours	Dependencies on other stakeholders

Table 7 takes some of the complexity metrics discussed and provides an example of a tool to measure the complexity of a weapons system development program during the planning process and when using complexity to develop SERs. The tool uses the metrics shown in Tables 3 and 4 and provides a menu for the PMO to assess complexity.

Table 7. Complexity Score Card

Complexity Assessment Tool						
Parameter	Low <25 pts	Medium <50 pts	High <75 pts	Very High <100	Weight	Total Complexity Index
Size	$<10M	$10–99M	$100–500M	>$500M		
Project Duration	<1 yr	<3 yr	<7 yr	>7 yr		
Ratio Budget/ Duration						
Organizational	PdM	PM	PM	PEO		
Budget	Yes	Some	Little	First Time		
Risk	Low	Med	High	Very High		
Technical Complexity	Low	Med	High	Very High		
Technological Maturity	Very High	High	Med	Low		
Dynamics	No	Little	Some	Yes		
Human Dimension	Component	Subsystem	System	SOS		
Number of Contr/Subs	<3	< 5	< 7	>7		
Software						
Total						

Complexity Leads to Delays

Table 8 shows the relationship of project complexity to the identified schedule delay factors. When more than one factor is present, they are listed in order of impact. Examination of Table 8 almost forces one to ask the question, *Which comes first, the complexity issue or the delay?* The answer to that question depends on the desired response. The complexity factors would be used to assess programs during the planning process to allow a for macro-level estimate using SERs. Similarly, the delay factors would also be used during the planning process as questions to be answered during the walk-through of the work breakdown structure. Together the elements provide a tool to be used during program execution.

Table 8. Combined Complexity and Delay Factors

Complexity Factors	Delay Factors
Structural	Competition at the prime contractor level
Sociopolitical, Dynamic	Administrative changes
Structural, Dynamic	Concurrency, overlap in time and effort between the development and production phases of a program
Uncertainty, Structural	Budget/funding delays, funding adequacy/stability
Uncertainty	Existence of prototyping
Structural	Separate contracts for each phase of the program
Structural	Priority of the program to the service relative to other ongoing programs
Structural	External guidance such as OSD or congressional direction, reviews, restrictions, and designations
Structural	Joint management with other agencies
Uncertainty	Technical difficulty
Uncertainty	Concept stability, or stability in mission, operational concepts, and doctrine
Uncertainty	Contractor delays
Dynamic	Delays due to contracting/contract negotiation/award delays
Uncertainty	External events such as inflation, earthquakes, labor strikes (force majeure)
Uncertainty	Major requirements stability, design freeze
Sociopolitical, Uncertainty, Dynamic	Program manager turnover
Uncertainty, Dynamic	Testing delays
Uncertainty, Dynamic	Rework
Uncertainty, Sociopolitical	External events such as inflation, earthquakes, labor strikes, etc.

Schedule Delays, Complexity, and Historical Learning

A development project or program is a dynamic system with feedback loops. Invariably, decisions taken to address one problem have an impact on or create new problems. We believe the schedule and complexity factors discussed in this paper can be effectively applied to the analysis and development and eventual execution of the schedule. Finally, an appreciation of the historical performance of development programs can and should be used to better inform the development of weapons system development schedules.

While the case for complexity as a significant schedule driver seems compelling, “complexity” is complex to define and difficult to measure. Further, a “Total Complexity Index” is appealing, but reducing a vector whose components are challenging to quantify to a scalar quantity is imposing.

There have been some interesting and valuable efforts to find observable proxies for complexity. For example, physical complexity as defined by the density of equipment within a platform (Grant, 2008; Terwilliger, 2015) has been studied as a cost driver. Likewise, “virtual” complexity, as measured perhaps by lines of software code, is very promising. However, these capture only a few of the total-complexity vector described in Table 8.

Summary and Concluding Comments

Our primary purpose in this effort has been to build a case for new empirical sources and methods for acquisition schedule estimation. We have tentatively identified CAQDAS software.

We got there by considering the inherent difficulties in schedule estimation: "incidental" factors, variation in outcomes due to program management decisions, and the incentives endemic to source selections that reward unrealistic estimates (cost, schedule, and performance). We reported evidence from the literature that supported both common and idiosyncratic schedule drivers across programs.

Support for the common-factors perspective comes from "meta-studies" of program outcomes that can be reduced to several major program themes (such as quality of communication, management competence, and degree of trust between the major players).

While these lines of inquiry are interesting and promising, defining, operationalizing, and measuring are difficult (at best) using methods within the current state of practice.

One useful next step in advancing the art and science of schedule estimation is new forms of data analysis. Fortunately, several tools have recently emerged for analyzing quantitative and qualitative data. We view developments in qualitative data analysis to be more promising—particularly concerning variables (such as quality of communication) that are difficult to map to real numbers.

References

Amara, J., & Franck, R. (2021). *The U.S. defense economy*. Cambridge University Press. https://www.cambridge.org/core/elements/us-defense-economy/51D8A8897B2985E3ACE6FC8CDCDED130

Atkinson, R. (1999). Project management: Cost, time, and quality, two best guesses and a phenomenon, it's time to accept other success criteria. *International Journal of Project Management*, *17*(6), 337–342. http://www.sciencedirect.com/science/article/pii/S0263786398000696

Baccarini, D. (1996). The concept of project complexity: A review. *International Journal of Project Management*, *14*(4), 201–204.

Bosch-Rekveldt, M., Jongkind, Y., Mooi, H., Bakker, H., & Verbraeck, A. (2011). Grasping project complexity in large engineering projects: The TOE (technical, organizational and environmental) framework. *International Journal of Project Management*, *29*(6), 728–739. https://www.doi.org/10.1016/j.ijproman.2010.07.008

Boyd, D. S., & Mundt, B. D. (1995). Schedule estimating relationships for the engineering and manufacturing of bomber, transport, tanker, and surveillance aircraft systems. *The Journal of Cost Analysis*, *12*(1), 131–154.

Chairman of the Joint Chiefs of Staff. (2018). *Manual for the operation of the Joint Capabilities Integration and Development System*. Office of the Under Secretary of Defense for Acquisition and Sustainment. https://www.acq.osd.mil/asda/jrac/docs/2018-JCIDS.pdf

Dao, B., Kermanshachi, S., Shane, J., & Anderson, S. (2016). Project complexity assessment and management tool. *Procedia Engineering*, 145, 491–496. https://www.doi.org/10.1016/j.proeng.2016.04.026

Drezner, J. A., & Smith, G. K. (1990). *An analysis of weapon system acquisition schedules* (R-3937-ACQ). RAND. https://www.rand.org/pubs/reports/R3937.html

Franck, R., Hildebrandt, G., & Udis, B. (2016). Toward realistic acquisition schedule estimates (SYM-AM-16-029). *Proceedings of the 13th Annual Acquisition Research Symposium*, 95–116. http://my.nps.edu/documents/105938399/108639550/SYM-AM-16-029.pdf/12c9563c-164f-495d-935f-8b4b969d31d2

GAO. (2010). *Strong leadership is key to planning and executing stable weapon programs* (GAO 10-522). https://www.gao.gov/products/gao-10-522

GAO. (2015). *Acquisition reform: DoD should streamline its decision-making process for weapon systems to reduce inefficiencies* (GAO-15-192). https://www.gao.gov/products/gao-15-192

Grant, B. P. (2008). *Density as a cost driver in naval submarine design and procurement* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/4085

Harmon, B. R., Ward, L. W., & Palmer, P. R. (1989). *Schedule estimating relationships for tactical aircraft, cost analysis applications of economics and operations research: Vol 332* (pp. 259–280).

Jardine, S., Moseley, M., Moul, J., & Trapp, D. (2019). Schedule estimating relationship (SER) development using missile and radar datasets. *Proceedings from 2019 ICEAA Professional Development & Training Workshop*. https://www.iceaaonline.com/ready/wp-content/uploads/2019/06/AM07-Paper-Schedule-Estimating-Relationshop-Development-Jardine.pdf
Jimenez, C. A. (2016). *Predicting schedule duration for defense acquisition programs: Program initiation to initial operational capability* [Thesis, Air Force Institute of Technology]. AFIT Scholar. https://scholar.afit.edu/cgi/viewcontent.cgi?article=1244&context=etd
Jimenez, C. A., White, E. D., Brown, G. E., Ritschel, J. D., Lucas, B. M., & Seibel, M. J. (2016). Using pre-Milestone B data to predict schedule duration for defense acquisition programs. *Journal of Cost Analysis and Parametrics*, *9*(2), 112–126. https://www.doi.org/10.1080/1941658x.2016.1201024
Kahneman, D. (2011). *Thinking fast and slow*. Farrar, Straus, & Giroux.
Light, T., Leonard, R. S., Pollak, J., Smith, M. L., & Wallace, A. (2017). *Quantifying cost and schedule uncertainty in major defense acquisition programs (MDAPs)* (RR-1723-AF). RAND. https://www.rand.org/pubs/research_reports/RR1723.html
Light, T., Leonard, R. S., Smith, M. L., Wallace, A., & Arena, M. V. (2018). *Benchmarking schedules for major defense acquisition programs* (RR-2144-AF). RAND. https://apps.dtic.mil/sti/citations/AD1085832
Magee, C. L., & de Weck, O. L. (2004). Complex system classification. *INCOSE International Symposium*, *14*(1), 471–488. https://www.doi.org/10.1002/j.2334-5837.2004.tb00510.x
Maier, M. (1995). Social systems engineering and architecture. *Proceedings of the Fifth International Symposium of the National Council on System Engineering*, *1*, 139–146. http://libsys.uah.edu/library/incose/Contents/Papers/95/95117.pdf
Mayer, K. R. (1993). *The development of the advanced medium range air-to-air missile: A case study of risk and reward in weapon system acquisition* (N-3620-AF). RAND. https://www.rand.org/pubs/notes/N3620.html
McNicol, D. L. (2015). *Are changes in acquisition policy and process and in funding climate associated with cost growth?* (IDA Document NS D-5448). Institute for Defense Analyses. https://www.ida.org/-/media/feature/publications/a/ar/are-changes-in-acquisition-policy-and-process-and-in-funding-climate-associated-with-cost-growth/d-5448.ashx
McNicol, D. L. (2020). *Program duration, funding climate and acquisition policy* (IDA Document NS D-13165). Institute for Defense Analyses. https://www.ida.org/research-and-publications/publications/all/p/pr/program-duration-funding-climate-and-acquisition-policy-conference-presentation
McNicol, D. L., & Wu, L. (2014). *Evidence of the effect of DoD acquisition policy and process on cost growth of major defense acquisition programs* (IDA Paper P-5126). Institute for Defense Analyses. https://apps.dtic.mil/sti/pdfs/ADA609472.pdf
O'Hanlon, M. (2021, March 5). *A conversation with HASC Chairman Adam Smith*. Brookings. https://www.brookings.edu/events/a-conversation-with-hasc-chairman-adam-smith/
O'Neil, W. D. (2011, July) Cost growth in major defense acquisitions: Is there a problem? Is there a solution? *Defense Acquisition Research Journal, 59*, 277–294. https://www.dau.edu/library/arj/_layouts/15/WopiFrame.aspx?sourcedoc=/library/arj/ARJ/arj59/Oneil_ARJ59.pdf&action=default
Peck, M. J., & Scherer, F. M. (1962). *The weapons acquisition process: An economic analysis*. Graduate School of Business Administration, Harvard University.
Pickar, C. K. (2018). Informing DoD program planning through the examination of the causes of delays in acquisition using acquisition data (SYM-AM-18-079). *Proceedings of the 15th Annual Acquisition Research Symposium*, 75–93. https://calhoun.nps.edu/bitstream/handle/10945/58762/SYM-AM-18-079-019_Pickar.pdf?sequence=1
Pickar, C., & Franck, R. (2019). Making time from data: Toward realistic acquisition schedule estimates (SYM-AM-19-062). *Proceedings of the 16th Annual Acquisition Research Symposium*, *II*, 29–50. https://calhoun.nps.edu/bitstream/handle/10945/62902/SYM-AM-19-062.pdf?sequence=1&isAllowed=y
Pinto, J. K. (2000). Understanding the role of politics in successful project management. *International Journal of Project Management*, *18*(2), 85–91. https://www.doi.org/10.1016/s0263-7863(98)00073-8

Riposo, J., McKernan, M., & Kaihoi Duran, C. (2014). *Prolonged cycle times and schedule growth in defense acquisition: A literature review* (RR-455-OSD). RAND and National Defense Research Institute. https://www.rand.org/pubs/research_reports/RR455.html

Sargut, G., & McGrath, R. G. (2011). Learning to live with complexity. *Harvard Business Review*, *89*(9), 68–76. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=21939129&retmode=ref&cmd=prlinks

Shackelford, M. D. (2021). *From EMD to Milestone C and beyond: Common issues that affect developmental programs transitioning into production* (Forum Paper No. 42). Mitchell Institute. https://mitchellaerospacepower.org/event/from-emd-to-milestone-c-and-beyond-common-issues-that-affect-developmental-programs-transitioning-into-production/

Sheard, S. A., & Mostashari, A. (2009a). A complexity typology for systems engineering. *Systems Engineering*. http://www.incose.org/flc/EVENTS/Documentation/Sheard_Complexity_Typology.pdf

Sheard, S. A., & Mostashari, A. (2009b). Principles of complex systems for systems engineering. *Systems Engineering*, *12*(4), 295–311. https://www.doi.org/10.1002/sys.20124

Sinha, S., Kumar, B., & Thomson, A. (2006). Measuring project complexity: A project manager's tool. *Architectural Engineering and Design Management*, *2*(3), 187–202. https://www.doi.org/10.1080/17452007.2006.9684615

Smith, G. K., & Friedman, E. T. (1980). *An analysis of weapon system acquisition intervals* (R-2605-DR&E/AF). RAND. https://www.rand.org/pubs/reports/R2605.html

Terwilliger, K. M. (2015). *Investigating outfitting density as a cost driver in submarine construction* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/47336

Thamhain, H. J. (2005). *Management of technology: Managing effectively in technology-intensive organizations*. John Wiley & Sons. http://books.google.com/books?id=kd5nAQAAQBAJ&hl=&source=gbs_api

Trudelle, R., White, E. D., Koschnick, C., Ritschel, J., & Lucas, B. (2017). Estimating an acquisition program's likelihood of staying within cost and schedule bounds. *Defense Acquisition Research Journal*, *24*(4), 600–624. https://www.proquest.com/openview/30484ba06490b3090a2df9673d4fc8bc/1?pq-origsite=gscholar&cbl=29577

Werner, B. (2019, December 5). Coast Guard to re-compete multi-million dollar offshore patrol cutter contract. *USNI News.* https://news.usni.org/2019/12/05/coast-guard-to-re-compete-multi-billion-dollar-offshore-patrol-cutter-contract

Whitty, S. J., & Maylor, H. (2009). And then came complex project management (revised). *International Journal of Project Management*, *27*(3), 304–310. https://www.doi.org/10.1016/j.ijproman.2008.03.004

Williams, T. (2002). *Modelling complex projects*. John Wiley & Sons Inc.

Williamson, O. (1996). *The mechanisms of governance*. Oxford University Press.

Technical Data Package Independent Assessment

Nicholas Haas—is a Group Leader, Acquisition Strategy & Management, at The MITRE Corporation with over 20 years of Public and Private sector service which includes Active-duty service with the United States Air Force and Reserve duty with the Alabama National Guard. He also has worked as a civilian with the United States Army Corps of Engineers, Army Contracting Command, and currently advises various Defense and Federal Agencies on Acquisition and Contracting Source Selection best practices and risk mitigation. He has a Bachelor of Science in Business Administrations from The Citadel along with 2 Master's degrees from the Florida Institute of Technology in Business Administration and Contracts/Acquisition Management. Also, he has 3 tours to Southwest Asia/Middle East as a Contingency Contracting officer. He is DAWIA Level III in Contracting and is currently the Commanding Officer for the Alabama National Guard's only Acquisition Team. [nhaas@mitre.org]

Virginia Wydler—is Chief Acquisition Scientist at The MITRE Corporation, supports various Intelligence Community and Defense sponsors with their acquisition strategies and source selections. She has more than 25 years of experience in acquisition and contracting, both federal government and commercial. She is a former Navy Contracting Officer for major systems acquisitions. She holds an M.S. in National Security Strategy, Industrial College of the Armed Forces; an M.S. in Acquisition and Contracting, Naval Postgraduate School; and B.S. in Business Administration, University of Maryland. She has published research papers and articles in such publications as the Defense Acquisition Journal and National Contract Management Association Journal. She is Certified Professional Contracts Manager (CPCM), and Fellow with National Contract Management Association (NCMA). She was DAWIA Level III Contracting while with Civil Service. [vwydler@mitre.org]

Raymond Eresman—The MITRE Corporation.

Kathleen Hyatt—The MITRE Corporation.

Scott Patton—The MITRE Corporation.

Dr. Beverly Stark-Kublin—The MITRE Corporation.

Abstract

The Program Management Office (PMO) is responsible for the quality and integrity of the data associated with system delivery and performance. When competing a new system or system update, the Program Manager (PM) is required to complete acquisition planning activities for a specific procurement and to develop a well-conceived acquisition strategy. This activity includes developing and validating Technical Data Package (TDP) requirements. Often, the program staff are committed to managing the current system and do not have the resources to fully review and validate the TDP for a major competition. Therefore, the TDP may not represent the best product to industry.

A PMO may consider having an outside entity conduct a review of their technical data to assess the readiness and viability of the TDP. An Independent Assessment (IA) may provide significant information for understanding what technical data is available, what data is missing for a competitive solicitation, and what needs to be purchased under a new or follow-on contract. An IA can assist with conducting the first steps in the competitive process of developing requirements and conducting market research. The purpose of an IA is to provide the PMO with additional validation that the TDP is sound for a solicitation and subsequent contract.

Executive Summary

The Program Office is required to comply with MIL-STD-31000B for Technical Data Packages. MIL-STD-31000B requires a technical description of an item adequate for supporting an acquisition, production, engineering, and logistics support. The TDP needs to provide an authoritative technical description of an item that is clear, complete, and accurate, and in a form and format adequate for its intended use.

A thorough analysis of technical data during requirements development and market research can have a positive impact on the rest of the source selection process. Releasing conflicting or missing technical data with the request for proposal (RFP) may result in poor or deficient proposals from industry. Having a strong assessment of the TDP will support a smoother approval process during the initial phase of the acquisition for Acquisition Strategy Panel (ASP) review, to mitigate questions or concerns about the readiness of the competitive package.

The IA may be conducted by an independent non-conflicted party outside of the program office and acquisition organization. The team should be comprised of both acquisition and technical subject matter experts (SMEs).

This research suggests how that independent party may conduct the assessment and provide feedback to the program office prior to RFP release. The independent assessment may be conducted in three phases of the review process:

- Initial Assessment (Phase I). The initial steps include research into TDP compliance documents, including military standards and federal agency guidance, to set the assessment parameters, and conducting an initial review of the TDP documentation. The IA should map the technical documentation against both federal and agency TDP Guidance, identifying redundancies across the various organizations. Once this analysis is complete, the team prepares questions for the interview sessions with the program functional groups representing the PMO.
- Functional Assessment (Phase II). In this phase, the IA team reviews the artifacts with each PMO functional group lead, compiling a recommended list of TDP data with links and locations for the individual documents. The IA team may assess the TDP artifacts by these functional group areas as outlined below. Organization and access plans for the voluminous data are important elements in a TDP review.
- Final Assessment, Comparative Analysis (Phase III). A comparison of the Program TDP with other major system acquisitions within the agency or other federal agencies for lessons learned and to address any omitted or conflicting documentation. This can include a list of applicable documents and include designations of compliance or reference. A peer review by another PMO that has recently conducted a competition can be invaluable in ensuring a quality RFP and source selection process.

The research paper recommended an in-depth review of the technical data by functional area, aligning the content with the PMO structure and Military Standard (MIL-STD) 31000B functional criteria. The PMO SMEs within these functional areas can provide the documentation for the IA. These areas can be adapted based on the PMO organization and its functional staff.

Purpose

A Program Office may consider having an outside entity conduct a review of their technical data to assess the readiness and viability of the Technical Data Package (TDP) for a new program start or a re-competition of a system with existing technical data. The Independent Assessment (IA) will serve as an additional data point for understanding what technical data is available, what data is missing for a competitive solicitation, and what needs to be purchased under a new or follow-on contract.

An independent assessment will assist with conducting the first step in the competitive process of developing requirements and conducting market research to attain industry interest.

A thorough analysis of the technical data during requirements development and market research can have a high impact on the rest of the process by mitigating risk of conflicting or missing technical data after the RFP is released, resulting in poor or deficient proposals from industry. Having strong assessment of the TDP will also support a smoother approval process during the Acquisition Strategy Panel (ASP) review, limiting questions or concerns about the readiness of the competitive package.

The Independent Assessment needs to be conducted by an independent non-conflicted party, outside of the program office and acquisition organization:

- a non-conflicted private concern familiar with the acquisition process and technical requirements development, typically a small business service company,
- a Federally Funded Research and Development Center,
- an academic organization, such as the Defense Acquisition University,
- a professional organization, chartered to conduct assessments.

Background

The Program Office is required to comply with MIL-STD-31000B for Technical Data Packages (DoD, 2018). MIL-STD-31000B requires a technical description of an item adequate for supporting an acquisition, production, engineering, and logistics support. The TDP needs to provide an authoritative technical description of an item that is clear, complete, and accurate, and in a form and format adequate for its intended use.

Technical data is defined as recorded information, regardless of the form or method of the recording, of a scientific or technical nature, including computer software documentation (DoD, 2018, p. 8). The term does not include computer software or data incidental to contract administration, such as financial or management information. (DFARS Clause 252.227-7013). Therefore, technical data encompasses a broad amount of documentation that can be available to develop and deliver a system.

The purpose of the TDP is to provide a technical description of an item that is clear, complete, and accurate and in a form and format adequate for its intended use. TDPs define the physical and functional characteristics of the accepted configuration of the item and its subordinate assemblies, subassemblies, and parts (DoD, 2018, p. 10).

The TDP comes under the umbrella of technical data management within defense weapons systems acquisition (OUSD[A&S], 2020). The Technical Data Management process provides a framework to acquire, manage, maintain, and ensure access to the technical data and computer software required to manage and support a system throughout the acquisition life cycle (DAU, 2021, Sec 4.3.2.4). Key technical data management considerations include understanding and protecting government intellectual property and data rights, achieving competition goals, maximizing options for product support, and enabling performance of downstream life-cycle functions. DoDI 5000.85, 3D.2.b.(5)(k) IP and 3D.3.c.(5) IP Strategy contains IP and IP Strategy policy for Major Capability Acquisition programs.

Acquiring the necessary data and data rights provide the ability to re-compete item acquisition, upgrades, and sustainment activities in the interest of achieving cost savings. The lack of technical data and/or data rights often makes it difficult or impossible to award contracts to anyone other than the original manufacturer, thereby taking away much or all of the government’s ability to reduce total ownership costs (DAU, n.d.).

Technical Data Package Review

The Independent Assessment of a program TDP results in a recommendation whether the technical artifacts are sufficient for the initial release of the RFP.

The information below details the TDP review process that should be followed to ensure a complete analysis of the state of the technical data for posting the Bidders Library for source selection. These task objectives should be included in any statement of work for the IA contractor to ensure a credible review and report.

- Review the TDP artifacts documented in the TDP documentation provided by agency functional area leads for readiness and sufficiency in a competitive source selection.
- Assess the technical data completeness for the artifacts to be posted to the Bidders' Library for use in the RFP phase of the source selection.
- Conduct an in-depth review of the data by functional area, aligning the content with the Agency Technical Data Package Guidance and Military Standard (MIL-STD) 31000B for TDP content.
- Assessment Program Office team's institutional knowledge of applicable federal and military documents to ensure compliance with regulations, guidance, and experience-based best practices.
- Included consideration of data rights, ownership of proprietary software, proprietary business practices, and other items that may/could limit the government's ability to publish key aspects of the program within their Bidders' Library. Ensure the program's contracting and legal offices are engaged in the review processes.

The IA should consider the following assumptions:

- The documents reviewed were provided to the IA team as of the effective date of the final data feed from the program office and technical leads.
- The assessment is based on documents supplied by the program office posted in the TDP artifact library or referenced in any TDP worksheet or spreadsheets.
- All documents provided for review are deemed legally sufficient and approved through legal vetting.
- The Bidders' Library may be divided into areas such as TDP content, governing DoD or agency directives, system documentation, and other information about the agency and the program, depending upon the program source selection team's preferences.
- A document is included in the library as part of the TDP section if it helps to provide a clear picture of the level of work expected of the bidding contractor in the performance of their duties. Documents outlining procedures, checklists, or data (e.g., defense design, map data) may be provided post-award.
- The acquisition may include both development and production system delivery for purposes of TDP definition.
- Top Secret/Sensitive Compartmented Information (TS/SCI) TDP documents may be necessary for a review. The RFI should discuss the need for TS/SCI clearances. The IA Team assumes that the government will have standard operating procedures (SOPs) to view TS/SCI documents during the RFP process.
- Source code is not reviewed during the IA process but is a consideration for the library.

The Final Report should include these categories of analysis and recommendations:

Observations are defined as something that the assessment team took note of throughout the assessment process and determined to be worthy of mentioning for a possible future acquisition.

Considerations are items that the assessment team noticed during the assessment and felt that the government could benefit from applying these points, but they are not necessarily strong enough to warrant a recommendation rating.

Recommendations are the IA team's guidance to the government on the viability and accuracy of the TDP to ensure that the information presented will benefit them throughout their acquisition process.

Assessment by Functional Area

Functional areas will most likely be aligned with the Program Office, rather than the system being acquired, because most technical staff own or maintain the technical data by their functional office code. They would be most knowledgeable about what data is available and version control.

Not all documents will have the proper classification and distribution markings. Prior to uploading into the Bidders' Library, the government functional team must ensure that all artifacts and documents are properly labeled.

Notional functional areas are shown below and listed in more detail in Appendix 1. These areas can be adapted based on the Program Office organization and its functional staff.

- Requirements & Design
- Development & Integration
- Modeling & Simulation (M&S)
- Anti-Tamper
- Cybersecurity
- Software
- Test
- Verification
- Training
- Transition & Installation
- Operations & Sustainment

The documents should be organized logically and reviewed for relevance and completeness. Are the documents current and accessible by a potential bidder? Is the data releasable to a potential bidder? Are the documents relevant—Does the library contain enough data for the bidder to make an informed proposal? The agency should also consider developing a repository of common government, DoD, and agency documents for consideration with pointers to where the most current references are located (e.g., the System Specification in the Program Technical Baseline Library). This will support the contracting office to develop the list of applicable compliance and reference documents used later in the Statement of Work. This will streamline the library building process and promote consistency between the acquisition phases of the competition.

Independent Assessment Process

The independent assessment can be divided into three phases, as the documentation analysis and review evolve, and is updated based on internal reviews and feedback.

Initial Assessment (Phase I)

The initial steps include research into TDP compliance documents, including military standards and agency guidance, to set the assessment parameters, and conduct an initial review of the TDP documentation received from the agency. The team starts by performing a mapping of the technical documentation against both the MILSTD3100B and any agency TDP Guidance and identifying redundancies across the various organizational tabs. Once this analysis is complete, the team prepares questions for the interview sessions with the program functional groups representing the Program Management Office (PMO).

The IA team conducts interviews during the initial period of the task. These interviews should include most of the agency functional group leads.

Interview Questions: Each functional group lead is provided a series of questions in advance to help facilitate the discussion. The following questions can be asked during the interview:

- What items are classified?
- Which documents are marked "proprietary" by the prime contractor, and of those documents that are marked proprietary, are they considered contract deliverables?
- Can we use the proprietary marked documents in the Bidders' Library? If so, to what extent?
- Are any of the items listed in the TBL workbook populated by the functional leads considered insufficient regarding the contract delivery instructions? Do they meet the terms and conditions of the contract? Were they delivered on time?
- Charters are listed. Are those charters going to be part of the Bidders' Library, and if so, why? Charters are typically established post-RFP but may be relevant if they outline the scope of work expected in execution of the proposed contract.
- Are all the documents listed in the TBL workbook considered to be complete and current? It should be clear if the artifacts will need to be periodically revised or resubmitted with updates as the contract and program matures.
- How is the government intending to package all this information for the RFP, and how will it be made available to the bidders in a competitive environment?

Data Categories: The agency functional managers should be asked to place their content into three categories of data: (1) Full access of unclassified non-proprietary data for all interested vendors which will be stored in the Bidders' Library, and ultimately listed in Appendix J of the RFP, (2) Classified material which is part of the Bidders' Library for the RFP stored in a secure site with access control to be determined by the government, and (3) Sensitive or proprietary technical data to be provided to winning contractor at the time of award such as mission sets, charters for working groups (WGs), general lists of data, and source code.

Data Markings: The functional group needs to make note of incumbent markings, restrictions, and proprietary rights claims. The government needs to be prepared for an internal government legal review to protect against challenges by the current prime contractor to the agency on markings of data ownership for artifacts included in the Bidders' Library. For example, source code may be a candidate for proprietary claims from the incumbent, which cannot be shared with other vendors if true and backed by a legal determination.

The artifact file names should be standardized to the TDP artifact master list descriptions. This will become an essential aspect during the solicitation phase of the acquisition, where the TDP will be released to industry in the Bidders' Library.

Functional Assessment (Phase II)

In this phase, the IA team utilizes the artifact spreadsheet with each PMO functional group lead, compiling a recommended list of TDP data with links and locations for the individual documents.

The PMO team provides a spreadsheet of artifacts organized in a similar manner to the structure of the program office, referred to in this appendix as functional areas. Note that this does not necessarily align with the functional or sub-system components that make up the Program system. The IA team may assess the TDP artifacts by these functional group areas.

A sample of the list of artifact files and how they should be depicted in a table to capture all the functional data in the artifact spreadsheet is shown in Appendix 2.

The IA team conducts an extensive review of the artifacts by searching and checking each link tied to every document. Some documents may not be found in the specified location. The IA team may conduct multiple rounds of updates with the PMO leadership team to locate and recover the missing TDP artifacts.

The Phase II report contains the artifacts recommended by the IA team for inclusion in the Bidders' Library. The team considers factors such as accessibility by a bidder, utility, and appropriateness of each document, and the presence of any limiting factors such as proprietary information contained within the artifacts. The team also highlights any areas for clarification and provides recommendations where appropriate.

A sample of the list of recommended artifacts for the Bidders Library is shown in Attachment 3.

It is recommended that prior to uploading into the Bidders' Library, the government team must ensure that all artifacts and documents are properly labeled.

Final Assessment, Comparative Analysis (Phase III)

It is recommended that a Technical Data Package Independent Assessment Report be delivered to the Program Office for review and discussion. Based on that discussion, additional actions may be requested by the PMO to revise and/or update the information.

The IA team can deliver an IA Technical Data Package After-Action Report to include the following updated documentation and further analysis:

- Updated functional technical data analysis to address additional information such as software documents and systems specifications.
- Revised or updated TDP Bidders Library List, based on functional changes.
- A comparison of the Program TDP with other major systems acquisitions within the agency or other federal agencies for lessons learned and to address any omitted or conflicting documentation. This can include a list of Applicable Documents included designations of compliance or reference. A peer review by another PMO that has recently conducted a competition can be invaluable in ensuring a quality RFP and source selection process.

Acquisition Considerations

The agency's need for technical data varies greatly from program to program with multiple factors contributing to each individual program's specific TDP needs. Many times, these factors are driven by the program acquisition and life-cycle support strategies. Factors such as the maturity of the program, the maturity of the program's system and system integrators, different conceptual design data for concept evaluations, or complete sets of detailed design data set points are all factors to be considered when deciding what a TDP should look like for each individual program.

Any upcoming competitive acquisition should consider the following areas as part of their acquisition strategy relative to technical data:

- Data Management Strategy
- Long-Term Strategy for continuous competition
- Intellectual Property Considerations
- Technical Data and Rights in Data
- Contract Data Requirements List (CDRLs) content
- Open Architecture Standards
- Acquisition Best Practices

While a complete TDP may be built in a specific way for one program, it could significantly vary in other programs within the same agency. Further, throughout the RFP process, the TDP may be updated due to feedback from prospective bidders as they request more information, ask questions regarding the TDP, respond to an RFI, or offer technical solutions that the government may incorporate into the RFP.

The key takeaway is that TDPs are not a "one-stop shop" and there is not a single "right" list to choose from when a program develops their own TDP.

The TDP is vital to the success of any competitive acquisition. This analysis is intended to support the critical milestones and events that will be the next steps in the acquisition process. Figure 1 shows the milestones for a competitive acquisition and where the TDP fits into the process and where the TDP can impact the process (in green) through award.

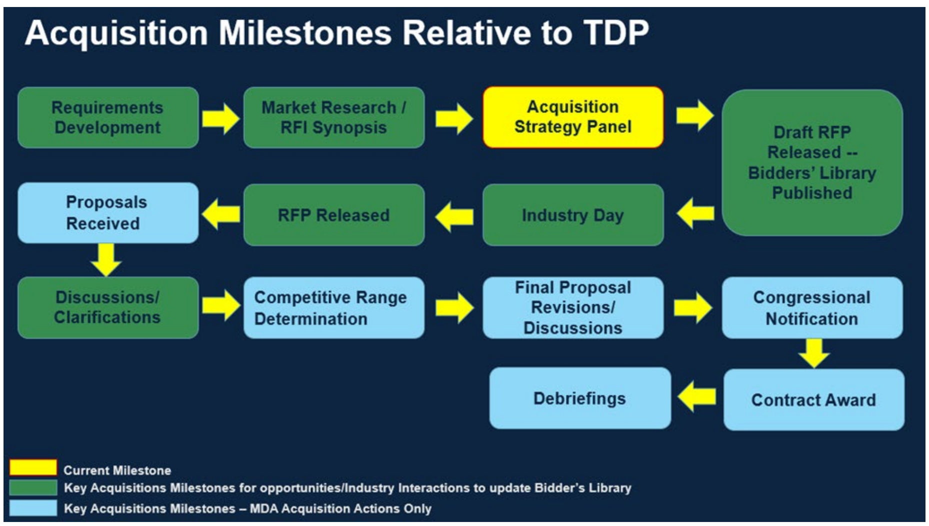


Figure 1. Acquisition Milestones Relatives to TDP

A further description of the impact of the TDP includes the following steps in the acquisition process as shown in Figure 1:

- Requirement development: TDP defined, validated, ready for the Bidders Library.
- Market research/RFI synopsis: Industry Q&A on technical data content and markings.
- Bidders' Library published: TDP available to industry, Q&A on content.
- Acquisition strategy panel: TDP readiness and data strategy addressed.
- Draft RFP released: TDP seen in full context of requirements, Q&A from industry.
- Industry Day: TDP addressed in a briefing, Q&A from industry.
- RFP released: Industry can have TDP Q&A during the solicitation period.

- Proposals received: TDP assumptions and conditions in proposal.
- Discussions/clarifications: TDP can be updated based on Q&A from industry
- Competitive range determination: TDP can impact score and ranking.
- Final proposal revisions/discussions: TDP updated based on Q&A from industry
- Contract award: TDP transferred to new contractor.
- Debriefings: any deficiencies may include TDP issues.

The TDP can be the topic of industry questions up through the final RFP release, allowing for updates and changes to the technical data. However, once the proposals are received, it is difficult to accommodate changes to the technical requirements since it may impact the competition's scope.

Conclusions and Recommendations

This research identifies an approach for a thorough review of the TDP to ensure that technical artifacts are sufficient for an acquisition and highlight areas that need improvement prior to release.

It is recommended that a PMO utilize an independent party to do a thorough review of the technical data prior to any major competition to assure the quality and integrity of the documentation to be utilized by industry to deliver a system.

The PMO should conduct an in-depth review of the technical data by functional area, aligning the content with the PMO structure and Military Standard (MIL-STD) 31000B functional criteria. The PMO SMEs within these functional areas can provide the documentation for the IA. These areas can be adapted based on the PMO organization and its functional staff.

The research concludes that it is appropriate to conduct a review of areas where the TDP may impact the acquisition strategy relative to technical data. These items may include the data management strategy, technical data and rights in technical data, and contract data requirements list (CDRL) content.

References

Defense Acquisition University. (n.d.). *Systems engineering brainbook: Technical data. Management.*
Defense Acquisition University. (2021, April). *Defense acquisition guidebook*.
Department of Defense. (2018, October). *Standard practice technical data packages* (MIL-STD 3100B).
Office of the Under Secretary of Defense for Acquisition and Sustainment. (2020, September 6). *The defense acquisition system* (DOD Directive 5000.01). Department of Defense. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/500001p.pdf

Appendix 1: Notional Functional Areas of Technical Data

Notional functional areas utilized for the TDP independent assessment are listed below and can be adapted based on the Program Office organization and its functional staff.

- Requirements & Design: The Requirements and Design functional area focuses on development of future program requirements as well as operations and sustainment of current fielded capabilities. The TDP artifacts listed in this section identify the major components and/or functional requirements that comprise the system.
- Development & Integration: This functional area contains artifacts relating to development and integration. A system must have the ability to creatively develop, mature, integrate, and test new technologies quickly and reliably to provide a state-of-the-art defense system. It must also respond to evolution in within the overall system's

sensors and weapons systems. This includes all aspects of the systems engineering process, including system design and algorithm software development.

- Modeling & Simulation (M&S): The modeling and simulation executable functional area focuses on those artifacts that will create an integrated digital approach across disciplines to support life-cycle engineering is in the planning stages. Models and Simulations will be implemented as a continuum across the Systems Engineering "V" as the authoritative sources of engineering data.
- Anti-Tamper: The Anti-Tamper (AT) functional area focuses on protecting the embedded Critical Program Information (CPI) and identifying design documentation which will enable exportability to foreign partners. The TDP artifacts provide systems security engineering (SSE) background/guidance to protect CPI and enable system exportability in accordance with applicable laws, regulations, policies, and procedures.
- Cybersecurity and Cyberspace Defense: Artifacts for the Cyber-resiliency function identifies cyber requirements in early development from current threat intelligence, complies with applicable statutes, regulations, the Risk Management Framework (National Institute of Standards and Technology [NIST] 800 series guidance), and conducts software assurance and cyber-resiliency testing throughout development, capability testing, ground tests, and continuous persistent cyber operations monitoring.
- System software and source code: These artifacts are developed, integrated, and tested for the future fielded capabilities. As an alternative, the TDP could include items such as Software Design Descriptions, Interface Design Descriptions, Algorithm Description Documents, operating manuals, and other documents that include descriptions of all hardware, software, firmware, middleware, hypervisors and other specialized application stacks, binaries, operating systems, and scripts/scripting engines, to include build matrices describing usage and platform associations. Including these items in the Bidders' Library would help the bidders understand the scope of the software. Documents should cover the areas such as the language, lines of code, number of subroutines, and interface requirements. What is needed is an understanding of the software and applications supporting mission critical functions, the planned development roadmap, and a listing of sample products to include Verification and Validation (V&V), Assessment and Authorization (A&A), software assurance requirements, scenario development, and any other information necessary for a bidder to understand the complexity and scope of work required. Consider an access plan for the bidders who do request to review the software. Otherwise, the actual software and source code files may be delivered post-award.
- Test: The test functional area includes the development, integration, and testing of system software with all additional hardware and external software required to achieve an integrated increment capability. Testing includes both element-level and system-level testing. System testing includes flight tests (FTs), digital predictive analysis, Hardware in the Loop (HWIL) Ground Tests (GTs), distributed GTs, and HWIL cybersecurity testing.
- Verification: This functional area identifies how the system must also respond to evolution in sensors and weapons systems of the overall system. As other elements of the system make changes or upgrades, the program must adjust accordingly to maintain critical integrated end-to-end capability. This includes all aspects of the systems engineering process including verification and validation as the resulting system must include a strong cyber security posture and overall system resiliency.
- Training: The training functional area seeks to safely separate test, evaluation, and training venues from real-world activities, and allow injection of high-fidelity simulations to run realistic scenarios on operational equipment and networks. The artifacts in this

area are necessary to maintain the operational capability of the requirement, to participate in exercises, to train, and to rehearse mission scenarios while the system is in an operational state or "on alert." The architecture will allow for scalable training over the operational architecture and will allow operators to train in their environment. Scalable training can vary from individual assets to regional capabilities, to the full global community.

- Transition & Installation: This functional area addresses continued development of system which will require the transition of software builds and the deployment of new hardware as required.
- Operations & Sustainment: The operations and sustainment functional area focuses on the ability to operate, maintain, and sustain the current globally deployed system while minimizing total ownership costs for current and future versions efficiently and effectively. The system will continue to support and interface with a variety of globally distributed sensors and communications elements hosted in a variety of facilities. The focus is to increase supportability and reduce hardware and software life-cycle costs of current and future variants in a technology environment that faces rapid turnover and requires increasing cyber resilience.

Appendix 2: Sample List of Technical Artifacts from PMO

A spreadsheet should be created that has columns for reference, title, description, version and/or date of the document, source where it can be found, OPR for the document, and any other notes that may be helpful. Additional columns may be added to identify Government Purpose Rights or Proprietary information, Classification, When to Release (e.g., pre-RFI Bidders' Library or post-Award), CDRL Reference, or other columns appropriate to the program. Locating the documents and verifying the sources can be an arduous task, and updated, clear, consistent record-keeping is essential. Using a spreadsheet allows for ease of maintenance and allows for sorting (to separate by functional area or classification for example).

Ref #	Title	Description	Version/ Date	Source	OPR	Notes
1	System Level Architecture Framework Documentation	Briefing outlining various system level architectural views	V4.7/23 April 2019	Agency Office Code *SharePoint Site link*)	CAG – John Smith	Includes both current and "to-be".
2	System Engineering Plan	Outlines the systems engineering processes within the program.	V9.3/4 June 2018	Agency Office Code *SharePoint Site link*)	SE/Jane Smith	Also include program-specific SEP
3	System Operator Manual	Outline of how to operate the current system.	V2.3/23 May 2017	Agency Office Code *SharePoint Site link*)	GMN/Sally Ride	Useful as reference for current operations.

Appendix 3: Sample Recommended Bidders' Library Content

	Functional Area (Example: Anti-Tamper)
1	Anti-Tamper Plan (Concept, Initial and Final) from Anti-Tamper Plan Template
2	Attack/Countermeasures Tree Analysis (in support of Anti-Tamper)
3	Technical report: study/services, anti-tamper plan
4	Technical report: study/services, attack countermeasure tree analysis

PAPERS ONLY

	Middle Tier Acquisition FY 2022 Budget Data Overview Amir Etemadi, George Washington University John Kamp, George Washington University
	Acquisition Warfare: A Proposal for a Unifying Concept Lieutenant Commander Ryan Hilger, Colorado State University
	Estimating Middle Tier Acquisition Schedule Risk John Kamp, George Washington University Amir Etemadi, George Washington University
	Build Back Better: The Reemergence of American Manufacturing Is Easier Said Than Done Learning from Building the Defense Industrial Base Symantha "Sam" Loflin, Defense Contract Management Agency
	The Price of Slavery: An Analysis of Human Trafficking Policy and Spending in Department of Defense Procurement Capt. Willis C. Crouch IV, U.S. Air Force 1st Lt. Austin L. Morris, U.S. Air Force 1st Lt. Kevin P. Peaslee, U.S. Air Force
	An Investigation of the Role of System Effectiveness in the Acquisition and Sustainment of U.S. Defense Systems: 1958 to 2021 John M. Green, Naval Postgraduate School

Middle Tier Acquisition FY 2022 Budget Data Overview

Dr. Amir Etemadi—Amir Etemadi received a Doctor of Philosophy degree in electrical engineering from the University of Toronto, Toronto, Canada. He is an Assistant Professor in the School of Engineering and Applied Science in the Department of Electrical and Computer Engineering and the Department of Engineering Management and Systems Engineering, and teaches masters- and doctorate-level courses in electrical engineering and engineering management. He is the Principal Investigator on several current research projects and has supervised over 50 masters and doctoral candidates for degrees in both electrical engineering and engineering management. [etemadi@gwu.edu]

Dr. John Kamp—John Kamp received a Doctor of Engineering degree in engineering management from the George Washington University. Kamp teaches masters-level courses in project and program management and is a faculty Graduate Research Advisor in the School of Engineering and Applied Sciences. He is a retired Naval Submarine Officer with extensive experience in research and development and program management. His research interests include engineering management, maritime systems, and acquisition system research. Kamp is a Fellow in the Royal Institution of Naval Architects and a member of several professional associations. [jckamp2018@gwu.edu]

Abstract

This research uses publicly-released data from 2018 to 2021, consisting of budget submissions, program-related reporting, and contemporaneous press releases, to describe how services took the same novel authorities and developed Middle Tier Acquisitions (MTAs) with differing structures, risks, and results to date. We acknowledge the cultural and personality differences, and concentrate on the different approaches to scoping project technical uncertainty and systemic complexity to fit within MTA constraints.

Research Issue Statement: This exploratory research examines MTA data from public data including budget documentation.

Research Results Statement: This research identifies significant trends associated with MTA application to date.

Keywords: Middle Tier Acquisition, Defense acquisition, innovation

Introduction

Congress recently created Middle Tier Acquisition (MTA) programs, which provide the military services rapid prototyping and fielding pathways with new program flexibilities and an explicit schedule constraint. The services are executing multiple MTAs, resulting in a set of MTA experiments related to development, execution, and governance. This paper summarizes MTA data extracted from fiscal year (FY) 2022 budget data and explores some features of MTA execution between services.

As this is exploratory analysis, we identify inferences that may be drawn from the project distribution and resource allocations in the Department of Defense (DoD) FY 2022 budget documentation and significant trends associated with MTA distributions and resource allocations.

Background

Congress enacted MTA processes in 2016, enabling processes to prototype or field new capabilities within 2 to 5 years of approval (National Defense Authorization Act [NDAA], 2015, sec. 804). Key statutory changes enabled service acquisition executives to bypass traditional requirements and acquisition processes and establish direct-reporting program managers for these rapid acquisition programs (NDAA, 2015). By 2019, the DoD had revised over two dozen acquisition-related directives, instructions, and memoranda,[76] and introduced two new acquisition paths—rapid prototyping and rapid fielding (Lord, 2019). In 2020, the DoD brought traditional acquisition, urgent acquisition, MTAs, software, business and services acquisitions into an Agile Acquisition Framework (Lord, 2020).

DoD rapid acquisition strategies typically have limited scope and objectives, senior leadership support and oversight, and process modifications removing obstacles to faster delivery (NDAA, 2015). Tate (2016) thought such processes also included using already mature or developed systems, in modular steps, with incremental production. The MTA schedule constraint resembles earlier acquisition innovations such as information systems acquisitions that emphasized commercial products and processes (Cha et al., 2014). Williams (2005) considered that poor defense program performance resulted from systemic failures, in particular when conventional program management approaches were used for complex, uncertain, and time-constrained programs.

The Government Accountability Office (GAO) is conducting significant research and analysis related to MTAs.[77] They provide a consistent perspective of DoD acquisitions. In 2019, they reported 35 MTAs started by the services by March 2019 (Oakley, 2019). We report 85 MTAs found in the FY 2022 budget documentation in the next section, summarized in Table 1.

Table 4. Middle Tier Acquisition Data Trend

Service	GAO (2019)	GAO 2020[a]	GAO 2021[b]	FY 2022 DoD Budget[c]
Air Force	24	8	11	39
Army	8	5	5	20
Navy	3	0	1	21
Other	0	0	0	5
Total	35	13	17	85

a – MTAs reviewed in GAO-20-439 (Oakley, 2020), b – MTAs reviewed in GAO-21-222 (Oakley, 2021), c – source: https://comptroller.defense.gov/Budget-Materials/Budget2022/

The GAO 2020 and 2021 reports provide substantial information on MTAs where planned costs exceed Major Defense Acquisition Program criteria (Major Defense Acquisition Program Defined, 2021). The GAO reports provide excellent summaries of selected MTAs and in-stride assessments of GAO concerns with MTA governance and execution.

Two papers related to MTAs are in the Naval Postgraduate School Defense Acquisition Innovation Repository.[78] Riel (2020) surveyed defense acquisition professionals and found schedule speed was perceived as less important than performance or cost. We reported on interim schedule modeling simulations seeded with GAO 2020 data (Etemadi & Kamp, 2021b). We defined schedule risk as the likelihood of exceeding a planned duration and showed that the MTA schedule risk to exceed 60 months is less than 0.2 (20%), and that MTAs with budgets larger than $1 billion are more likely to exceed 60 months (Etemadi & Kamp, 2021b).

MTA projects are executed within the defense market and defined by the number of competent sellers[79] and the number of entities setting product requirements (Etemadi & Kamp,

[76] These may be found at https://www.esd.whs.mil/Directives/issuances/dodd/.
[77] MTAs are treated as an acquisition reform by the GAO (Oakley, 2019).
[78] An extensive collection of defense acquisition research (Naval Postgraduate School, 2021).
[79] This number reflects the market competition; in the DoD market there are often few competent sellers, and the market is described as an oligopoly.

2021a). FitzGerald et al. (2016) described market segments by products (namely military-unique, military-adapted, and commercial systems) and whether market competition was constrained or viable. Chesbrough (2003) characterized corporate innovation models as open or closed, where closed innovation occurs inside the company, and open innovation includes external participation; Zoe Stanley-Lockman (2021) extends this model to DoD innovation, where traditional acquisition programs behave much like closed innovation systems. Following their reasoning, MTAs are not restricted to closed or open innovation systems, but should benefit from open innovation approaches, adaption of existing available and commercial systems, and a specific buyer setting requirements.

We used data from publicly available budget documentation. This paper summarizes the MTA projects within service and agency research, development, test, and evaluation (RDT&E) documentation and includes data for five instances of procurement funding supporting MTAs.

Findings

Table 5. FY 2022 Program Elements With One or More MTA Labels

BA	Line	PE.BLI	PE.Name	ORG	BA	Line	PE.BLI	PE.Name	ORG	BA	Line	PE.BLI	PE.Name	ORG
04	43	0604033F	Hypersonics Prototyping	AF	04	52	0603619A	Landmine Warfare and Barrier - Ac	ARMY	04	36	0603502N	Surface and Shallow Water Mine (	NAVY
04	48	0604327F	Hard and Deeply Buried Target De	AF	04	53	0603639A	Tank and Medium Caliber Ammun	ARMY	04	58	0603635M	Marine Corps Ground Combat/Sup	NAVY
04	53	0207100F	Light Attack Armed Reconnaissanc	AF	04	60	0603801A	Aviation - Adv Dev	ARMY	04	92	0604659N	Precision Strike Weapons Develop	NAVY
04	55	0207455F	Three Dimensional Long-Range Ra	AF	04	69	0604037A	Tactical Intel Targeting Access Noc	ARMY	04	95	0605512N	MEDIUM UNMANNED SURFACE VE	NAVY
04	67	1203164F	NAVSTAR Global Positioning Syste	AF	04	72	0604113A	Future Tactical Unmanned Aircraft	ARMY	04	99	0605518N	CONVENTIONAL PROMPT STRIKE (	NAVY
04	70	1206425F	Space Situation Awareness Systen	AF	04	73	0604114A	Lower Tier Air Missile Defense (LT	ARMY					
04	74	1206760F	Protected Tactical Enterprise Serv	AF	04	81	0604403A	Future Interceptor	ARMY					
04	75	1206761F	Protected Tactical Service (PTS)	AF										
04	76	1206855F	Evolved Strategic SATCOM (ESS)	AF										
BA	Line	PE.BLI	PE.Name	ORG	BA	Line	PE.BLI	PE.Name	ORG	BA	Line	PE.BLI	PE.Name	ORG
05	121	1206442F	Next Generation OPIR	AF	05	91	0604601A	Infantry Support Weapons	ARMY	05	121	0604282N	Next Generation Jammer (NGJ) In	NAVY
					05	94	0604622A	Family of Heavy Tactical Vehicles	ARMY	05	125	0604366N	Standard Missile Improvements	NAVY
					05	97	0604645A	Armored Systems Modernization (	ARMY	05	140	0604601N	Mine Development	NAVY
					05	101	0604741A	Air Defense Command, Control an	ARMY	05	160	0605215N	Mission Planning	NAVY
					05	108	0604802A	Weapons and Munitions - Eng Dev	ARMY	05	161	0605217N	Common Avionics	NAVY
					05	109	0604804A	Logistics and Engineer Equipment	ARMY	05	174	0304785N	ISR & Info Operations	NAVY
					05	113	0604818A	Army Tactical Command & Control	ARMY					
					05	132	0605042A	Tactical Network Radio Systems (L	ARMY					
					05	136	0605052A	Indirect Fire Protection Capability	ARMY					
					05	137	0605053A	Ground Robotics	ARMY					
					05	142	0605148A	Tactical Intel Targeting Access Noc	ARMY					
					05	148	0605232A	Hypersonics EMD	ARMY					
					05	153	0605625A	Manned Ground Vehicle	ARMY					
BA	Line	PE.BLI	PE.Name	ORG	BA	Line	PE.BLI	PE.Name	ORG	BA	Line	PE.BLI	PE.Name	ORG
07	167	0101113F	B-52 Squadrons	AF	07	208	0203743A	155mm Self-Propelled Howitzer Ir	ARMY	07	201	0605520M	MARINE CORPS AIR DEFENSE WEA	NAVY
07	177	0102326F	Region/Sector Operation Control (	AF						07	205	0101226N	Submarine Acoustic Warfare Deve	NAVY
07	183	0207040F	Multi-Platform Electronic Warfare	AF						07	210	0204311N	Integrated Surveillance System	NAVY
07	188	0207138F	F-22A Squadrons	AF						07	221	0206313M	Marine Corps Communications Sy	NAVY
07	202	0207417F	Airborne Warning and Control Sys	AF						07	223	0206623M	Marine Corps Ground Combat/Sup	NAVY
07	205	0207431F	Combat Air Intelligence System A	AF										
07	239	0302015F	E-4B National Airborne Operation	AF										
07	240	0303131F	Minimum Essential Emergency Co	AF										
07	246	0304260F	Airborne SIGINT Enterprise	AF										
07	250	0305015F	C2 Air Operations Suite - C2 Info S	AF										
07	267	0305206F	Airborne Reconnaissance Systems	AF										
BA	Line	PE.BLI	PE.Name	ORG	BA	Line	PE.BLI	PE.Name	ORG	BA	Line	PE.BLI	PE.Name	ORG
08	318	0608410F	Air & Space Operations Center (A(	AF						06	191	0605873M	Marine Corps Program Wide Supp	NAVY
										06	194	0305327N	Insider Threat	NAVY

Table 2 displays RDT&E program elements (Pes) with MTA projects. The columns reflect the service (Left = Air Force, Middle = Army, Right = Navy). The rows are grouped by Budget Activity (BA). The first group (BA 04 = Advanced Technology Development) has significant activity by all services. The Army has the most activity in the second group (BA 05 = Advanced Component Development and Prototypes), but the Air Force has the largest budgeted projects in this group. The Air Force has the most in the third group (BA 07 = Operational System Development) projects. The last group includes Air Force software factory projects (BA 08 = Software and Digital Technology Pilot Programs) and two Navy projects (BA 06 = RDT&E Management Support).

Table 6. Air Force 2022 MTA Summary

BA	Line	PE.BLI	MTA.Name	GAO.21.page	MTA.Start	MTA.End	Duration	Modular	Agile	FY2020	FY2021	FY2022	Type	Type.MTA
04	43	0604033F	ARRW	121	May-18	Mar-23	58	0	1	286000	386157	238262	MSL	RP
04	48	0604327F	M-Code/EAJ Developme		Oct-20	Sep-21	11	0	0	0	2150	0	MSL	RP
04	53	0207100F	Light Attack Armed aircı		Oct-20	Sep-21	11	0	0	1982	0	0	AIR	RP
04	55	0207455F	3DELRR		Jan-20	Dec-22	35	0	1	22469	19321	0	C3I	RP
04	67	1203164F	MGUE2	133	Nov-20	Sep-25	58	0	0	308215	0	0	SPACE	RP
04	3	1203164SF	MGUE2	133	Dec-20	Sep-25	57	0	0	0	205923	281191	SPACE	RP
04	70	1206425F	Deep Space Advanced F		Jan-22	Mar-25	38	0	0	29013	0	0	SPACE	RP
04	7	1206425SF	Deep Space Advanced F		Jan-22	Mar-25	38	0	0	0	33359	123262	SPACE	RP
04	74	1206760F	PTES	137	Nov-18	Dec-21	37	0	0	101583	0	0	SPACE	RP
04	75	1206761F	PTS	139	Jun-19	Jun-26	84	1	0	154237	0	0	SPACE	RP
04	12	1206761SF	PTS	139	Sep-20	Jun-24	45	1	0	0	200178	243285	SPACE	RP
04	76	1206855F	Evolved Stra	125	Sep-20	Sep-25	60	1	0	161882	0	0	SPACE	RP
04	13	1206855SF	Evolved Stra	126	Sep-20	Sep-25	60	1	0	0	71395	160056	SPACE	RP
05	121	1206442F	OPIR	135	Oct-18	Oct-23	60	0	1	1470278	0	0	SPACE	RP
05	22	1206442SF	Next-Gen O	135	Oct-18	Oct-23	60	0	1		11128900	1137393	SPACE	RP
05	22	1206442SF	Next-Gen O	135	Oct-18	Oct-26	96	0	1	0	482013	661098	SPACE	RP
05	7	1206442SF	FORGE	131	Sep-20	Sep-24	48	1	1		498283	514577	SPACE	RP
07	34	1203001SF	Force Element Termina		Feb-19	Mar-24	61	1	1	0	156736	98979	C3I	RP
07	167	0101113F	CERP (RVP)	123	Sep-18	Apr-22	43	1	0	175359	273020	484068	AIR	RP
07	167	0101113F	CERP Rapid Physical Pro		Apr-22	Jun-25	38	1	0	0	0	0	AIR	RP
07	177	0102326F	NCR-IADS		Apr-21	Jun-22	14	0	1	0	4795	0	C3I	RP
07	183	0207040F	Spectrum Warfare Attac		Oct-22	Jan-23	3	1	0	0	0	36607	C3I	RP
07	188	0207138F	F-22 Capabi	129	Sep-18	Sep-21	36	1	1	537232	663825	647296	AIR	RP
07	188	0207138F	Sensor Systems		Jun-22	Dec-26	54	1	1	75685	260921	262972	AIR	RP
07	188	0207138F	Navigation Systems		Oct-19	Sep-26	83	1	1	5224	9000	25540	AIR	RP
07	188	0207138F	Communication System		Oct-19	Sep-26	83	1	1	0	0	131270	AIR	RP
07	202	0207417F	AWACS		Oct-19	Sep-22	35	1	1	67341	123925	171014	AIR	RP
07	239	0302015F	Survivable SHF		Oct-19	Jun-24	56	0	0	24583	3462	25581	AIR	RP
07	240	0303131F	CVR Inc 2		Jul-21	Sep-26	62	1	0	12067	22284	0	C3I	RP
07	240	0303131F	Global ASNT Inc 2		Jul-21	Jun-25	47	1	0	117	21391	19729	C3I	RP
07	246	0304260F	Common SIGINT Develo		Oct-20	Sep-22	23	0	0	85157	127832	97546	C3I	RP
07	250	0305015F	C2AOS-C2IS modificatio		Oct-19	Sep-20	11	0	1	5206	0	0	C3I	RP
07	267	0305206F	Next Generation Senso		Jan-21	Sep-22	20	1	0	17338	54841	30198	AIR	RP
08	318	0608410F	AOC.WS	119	Jul-19	Jun-24	59	1	1	0	0	186915	C3I	RP
01	57	3010F	F-15EX	127	Mar-20	Jun-23	39	0	0	621100	1367147	1334822	AIR	RF
04	20	3010F	LAA		Jul-18	Sep-22	50	0	0	30000	0	0	AIR	RP
05	32	3010F	Link-16		Jun-21	Oct-25	52	0	0	46031	153083	52702	AIR	RF
05	33	3010F	Sensor Enhancements (		Jun-20	Jun-23	36	0	0	49002	122283	196825	AIR	RF
05	38	3010F	Rapid Global Mobility		Oct-18	Sep-22	47	1	0	3617	1106	100	AIR	RP

Note that the Air Force reported three Rapid Fielding MTAs (F-15EX, Link-16, and Sensor Enhancements). The largest budget items are space-related (OPIR, F-15EX procurement, or F-22 Capability Pipeline). Some budget reporting (OPIR, for example) does not provide a project end or transition at 60 months. Note that the Air Force is planning to retire the F-22 fleet “by the 2030 timeframe” (Insinna, 2021).

Table 7. Army 2022 MTA Summary

BA	Line	PE.BLI	MTA.Name	GAO.21.page	MTA.Start	MTA.End	Duration	Modular	Agile	FY2020	FY2021	FY2022	Type	Type.MTA
04	52	0603619A	Area Denial Capability		Mar-22	Mar-25	36	1	0	0	4995	34761	GND	RP
04	53	0603639A	Advanced Armor-Piercir		Oct-18	Mar-24	65	1	0	8572	0	0	GND	RP
04	60	0603801A	FLRAA Virtual Prototype		Aug-22	Mar-24	19	1	0	0	0	102648	AIR	RP
04	69	0604037A	TITAN		Sep-21	Jun-23	21	0	0	0	0	28347	C3I	RP
04	72	0604113A	FTUAS		Sep-22	Jun-25	33	1	1	0	33758	48197	AIR	RP
04	73	0604114A	LTAMDS	161	Oct-19	Sep-22	35	0	0	364154	308805	327690	C3I	RP
05	91	0604601A	NGSW-FC program		Apr-20	Sep-21	17	1	0	14095	9782	11107	GND	RP
05	94	0604622A	Leader Follower		Oct-21	Sep-25	47	1	0	4294	10249	21918	GND	RP
05	97	0604645A	Mobile Prot	163	Dec-19	Jun-22	30	0	0	273433	123992	137256	GND	RP
05	98	0604710A	IVAS	159	Nov-19	Apr-21	17	1	1	60599	7495	4934	GND	RP
05	108	0604802A	Precision Munition (Sni		Oct-21	Sep-23	23	0	0			9275	GND	RP
05	108	0604802A	Small Caliber Ammo fo		Oct-18	Jun-23	56	0	0	17432	26483	28372	GND	RP
05	113	0604818A	Unified Network Opera		Apr-19	Jun-21	26	0	1	3499	3522	3366	C3I	RP
05	132	0605042A	Integrated Tactical Net		Jan-21	Mar-26	62	1	0	22411	9754	17762	C3I	RP
05	136	0605052A	Enduring IFPC Inc 2		Jan-21	Sep-23	32	0	0	186369	153362	233512	C3I	RP
05	137	0605053A	Small Multipurpose Eq		Jul-19	Sep-21	26	1	0	8768	28555	29448	GND	RP
05	142	0605148A	TITAN		Jul-21	Sep-24	38	0	0	0	0	28347	C3I	RP
05	148	0605232A	LRHW		Oct-22	Sep-24	23	0	0	0	0	111473	MSL	RP
05	153	0605625A	OMFV	165	Jul-21	Sep-24	38	1	0	197304	171890	225106	GND	RP
07	208	0203743A	ERCA Incren	157	Jul-19	Sep-23	50	0	1	191076	217959	213281	GND	RP

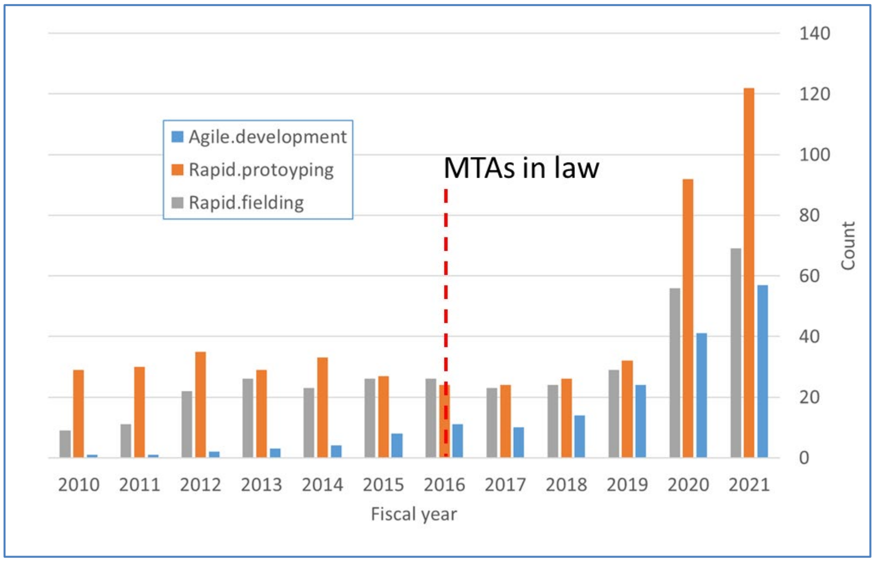
Table 8. Navy 2022 MTA Summary

BA	Line	PE.BLI	MTA.Name	GAO.21.page	MTA.Start	MTA.End	Duration	Modular	Agile	FY2020	FY2021	FY2022	Type	Type.MTA
04	36	0603502N	Medium Unmanned Su		Jul-20	Jun-27	83	1	0	22964	0	0	SHIP	RP
04	58	0603635M	Armored Reconnaissan		Jul-21	Sep-22	14	0	0	7465	17599	48563	GND	RP
04	59	0603654N	Expeditionary Diving Sy		Oct-19	Sep-25	71	1	0	911	1765	822	SHIP	RP
04	78	0604028N /	LIONFISH SUUV		Oct-19	Sep-22	35	0	0	0	4577	15881	SHIP	RP
04	92	0604659N	Convention	209	Oct-19	Jun-23	44	0	0	502435	0	0	MSL	RP
04	95	0605512N	Medium Unmanned Su		Jan-21	Sep-22	20	1	0	5200	3200	3500	SHIP	RP
04	99	0605518N	CPS prototy	209	Oct-19	Jun-23	44	0	0	0	766637	1372340	MSL	RP
05	125	0604366N	SM-2 Block IIIC		Oct-19	Sep-22	35	0	0	69180	56144	33412	MSL	RP
05	140	0604601N	Encapsulated Effector (		Oct-19	Sep-22	35	0	0	0	27000	40300	SHIP	RP
05	160	0605215N	Next Generation Naval		Oct-19	Sep-22	35	1	1	25420	35500	37606	C3I	RP
05	160	0605215N	Standardized Tester of		Oct-19	Apr-22	30	1	0	12975	14546	17772	C3I	RP
05	161	0605217N	MAGTF Agile Networkin		Jan-21	Apr-22	15	1	1	0	21133	18872	AIR	RP
05	174	0304785N	Integrated Communicat		Dec-19	Sep-22	33	1	1	8300	6095	1548	C3I	RP
06	191	0605873M	Marine Corps Wargami		May-19	Sep-22	40	0	1	11027	15000	23518	C3I	RP
06	194	0305327N	Counter Insider Threat		Oct-19	Sep-22	35	0	0	2592	2293	2581	C3I	RP
07	201	0605520M	Medium Range Interce		Jun-20	Sep-22	27	0	0	15300	52400	7800	MSL	RP
07	205	0101226N	Compact Rapid Attack V		Oct-21	Sep-26	59	0	0	0	13363	44854	C3I	RP
07	210	0204311N	Deployable Surveillanc		Oct-19	Sep-23	47	1	0	8500	26385	16592	C3I	RP
07	221	0206313M	Air Battle Management		Oct-19	Jun-22	32	1	1	6164	1290	1204	C3I	RP
07	223	0206623M	MEGFoS		Jun-20	Jun-22	24	1	1	3922	5753	12934	C3I	RP
07	223	0206623M	WSATCOM MCWS-X		Mar-21	Oct-21	7	1	1	20432	200	0	C3I	RF

Table 9. Other DoD/Agency 2022 MTA Summary

SVC	BA	Line	PE.BLI	MTA.Name	GAO.21.page	MTA.Start	MTA.End	Duration	Modular	Agile	FY2020	FY2021	FY2022	Type	Type.MTA
DOD	05	131	0604384BP	Rapid Opioid Counterm		Oct-19	Jun-22	32	1	1	13297	8417	11380	GND	RP
SOCOM	07	264	1160431BB	Weapons		Jan-20	Sep-23	44	1	0	1509	1604	1514	GND	RP
SOCOM	07	264	1160431BB	C-UAS		Mar-20	Sep-22	30	1	0	9671	5796	5195	GND	RP
SOCOM	07	264	1160431BB	Ground Organic Precisi(		Oct-19	Sep-26	83	1	0	7989	2290	15963	GND	RP
SOCOM	07	268	1160483BB	SOF Combat Diving (CBE		Dec-19	Nov-25	71	1	0	2580	2161	3183	SHIP	RP

Figure 1 shows the use frequency of terms related to MTA type programs.



.

Figure 11. MTA-Related Term Use Frequency

Figure 1 includes data before FY 2022 to show the historical term usage and the delay between MTA establishment in 2016 and use.[80] The number of rapid prototyping and fielding mentions in budget documents grew in FY 2020 and FY 2021, consistent with the increasing use of MTA authorities.[81] Figure 2 shows the distribution of FY 2022 RDT&E Pes with MTA labels[82] sorted by BA and service.

[80] See GAO-19-439 (Oakley, 2019).
[81] We did not count the FY 2022 usage trends.
[82] The values in Figure 1 are term use counts and, in Figure 2, counts of MTAs.

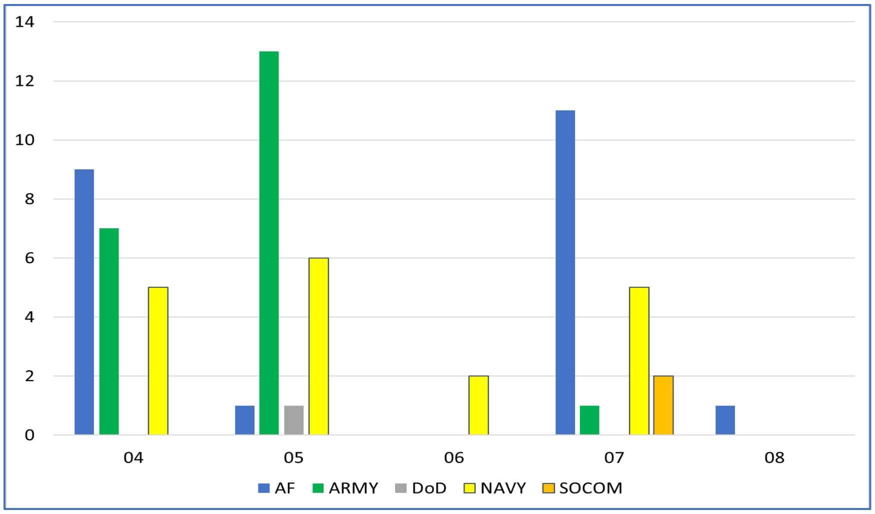


Figure 12. FY 2022 MTA Count by BA and Service

In FY 2022, the Army, Navy, and Air Force all had activity in BA 04 (Advanced Component Development and Prototypes), BA 05 (System Development and Demonstration), and BA 07 (Operational System Development). The distribution shows the Army leading new system development counts, while the Air Force was pushing both early development and operational systems. Figure 3 shows the same data sorted by service and commodity type.

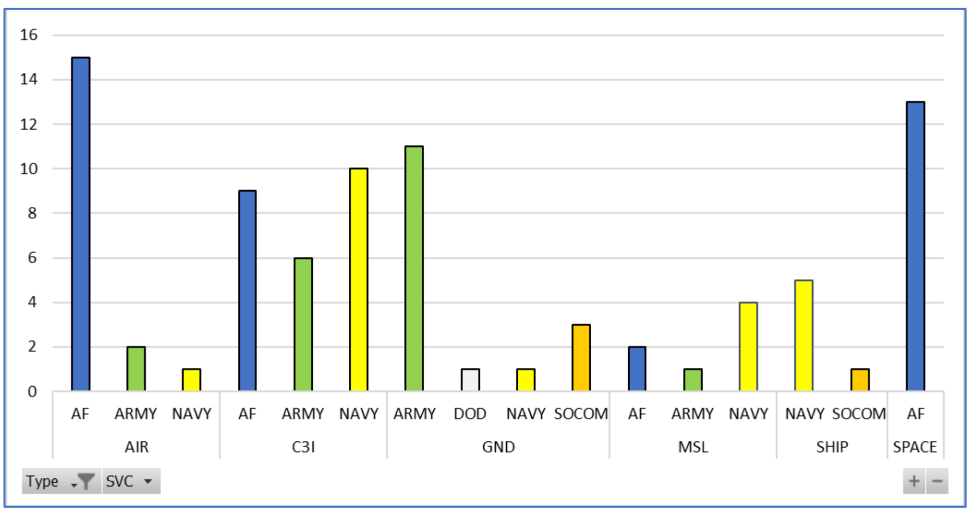


Figure 13. FY 2022 MTA Distribution by Service and Commodity Type

Figure 3 shows the Air Force emphasizing Air and space commodities, the Army emphasizing ground systems, and all three services investing in command, control, communications, and intelligence (C3I) projects. The C3I activity is consistent with use or adaptation of commercial products and processes. The Air Force activity includes projects transferred to Space Force. We present the resource allocations between FY 2020 and FY 2022 inclusive to highlight service trends. Figure 4 shows the spend for PEs with modularity labels.

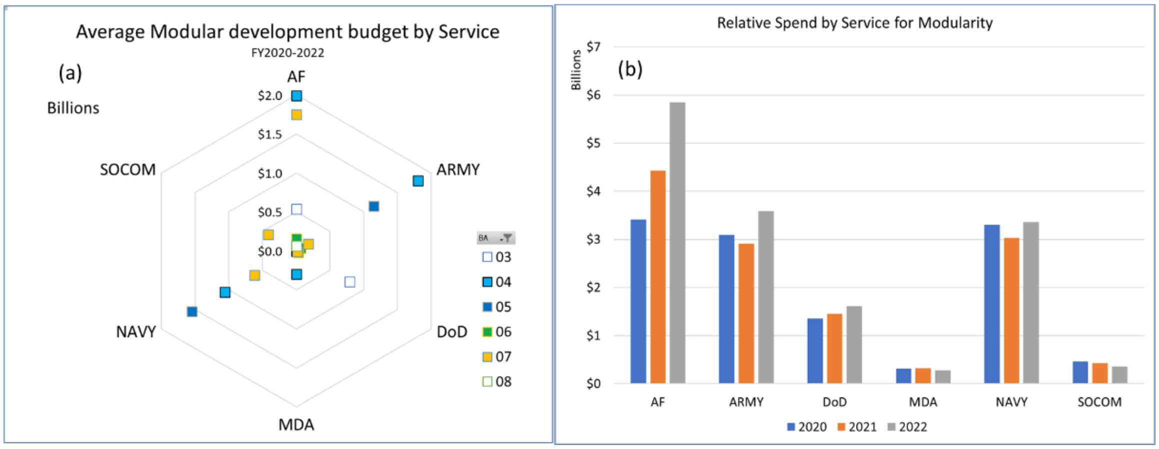


Figure 14. Resource Allocation Related to Modular Development

Figure 4a shows significant average service investment for all services related to modularity/modular development in BAs 04, 05, and 07. In Figure 4b, the Air Force shows an increasing trend, while the other services are relatively constant. Table 7 summarizes FY 2022 MTA modularity median duration and average budget median by commodity type.

Table 10. FY 2022 MTA Modularity Data Summary by Commodity Type

Type		Modular	Not Modular	Type		Modular	Not Modular
AIR	Duration	37	44.5	SHIP	Duration	71	35
	AVG budget	34171	50907		AVG budget	3304	14626
	Count	12	6		Count	4	2
C3I	Duration	35	33.5	SPACE	Duration	60	57.5
	AVG budget	13746	11690		AVG budget	77150	132555
	Count	13	12		Count	5	8
GND	Duration	36	30				
	AVG budget	11661	24542				
	Count	11	5				
MSL	Duration	*	35	Overall	Duration	43	36.5
	AVG budget	*	52912		AVG budget	13746	29514
	Count	0	7		Count	45	40

Table 7 shows the relative high cost and schedule risk of space projects. Modular MTAs have a longer median duration, but only the median average PE budgets are statistically different[83] ($\alpha = 0.1$). The ship MTA projects show long median durations due to schedule completions not being reported but shown as continuing. Modularity is being used to improve sustainment and supportability of operational or in-service systems or to create the ability to

[83] Mann-Whitney test, W-value = 1737, p-value = 0.082.

insert future upgrades to systems faster or at a lower cost or risk. Figure 5 shows the resource allocation to Agile projects.

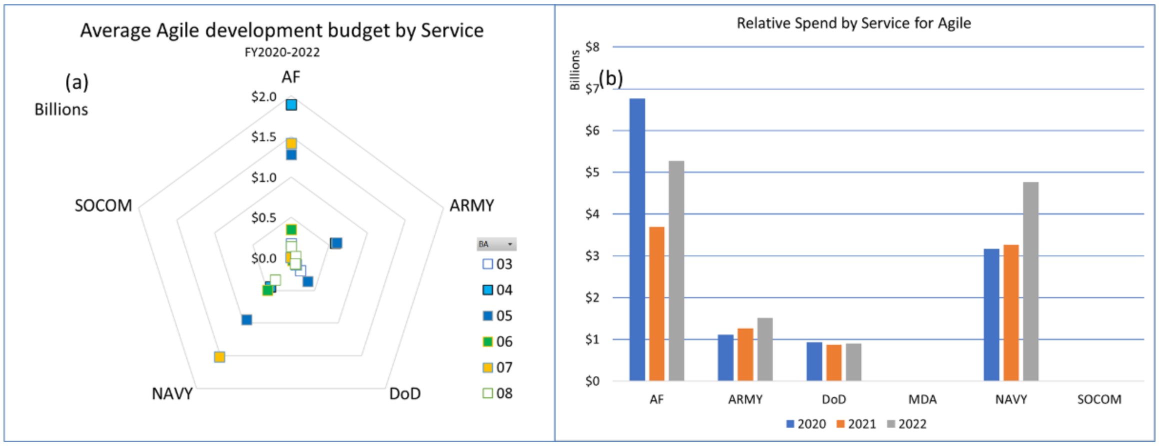


Figure 15. Resource Allocation Related to Agile Development

Figure 5 shows that the Air Force and Navy are making significant investment in Agile projects. Note that both the Navy and Air Force allocated significant BA 07 (Operational Systems Development) to PEs with Agile-related MTAs. Table 8 summarizes FY 2022 MTA Agile median duration and average budget median by commodity type.

Table 11. FY 2022 MTA Agile Data Summary by Commodity Type

Type		Agile	Not Agile	Type		Agile	Not Agile
AIR	Duration	36	39	SHIP	Duration	*	53
	AVG budget	43757	34126		AVG budget	*	5393
	Count	7	11		Count	0	6
C3I	Duration	32.5	35	SPACE	Duration	60	57
	AVG budget	7207	15098		AVG budget	498261	53961
	Count	12	13		Count	4	9
GND	Duration	32	36				
	AVG budget	24343	12154				
	Count	3	13				
MSL	Duration	58	31	Overall	Duration	35	38
	AVG budget	303473	45035		AVG budget	27318	18641
	Count	1	6		Count	27	58

Table 8 shows relatively few MTAs overall are engaged in Agile activities, with similar median durations; Agile MTAs have larger median average budgets, but the difference is not significant[84] (α = 0.1). Operational system software certification and approval processes may be reducing Agile use. Figure 6 shows the distribution of MTAs in the FY 2022 data associated with modular or Agile development.

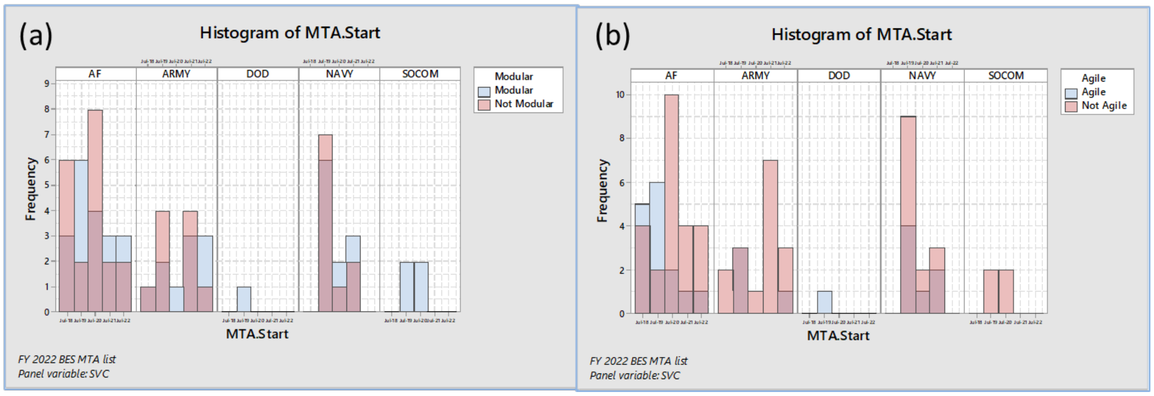


Figure 16. FY 2022 MTA Projects With Modular or Agile Labels by Start Date

Figure 6 shows marginal steady (marginal) to decreasing (Agile) use trends over time. More recent projects are more likely to not be identified as using Agile processes. Figure 7 summarizes MTA resource allocations by service.

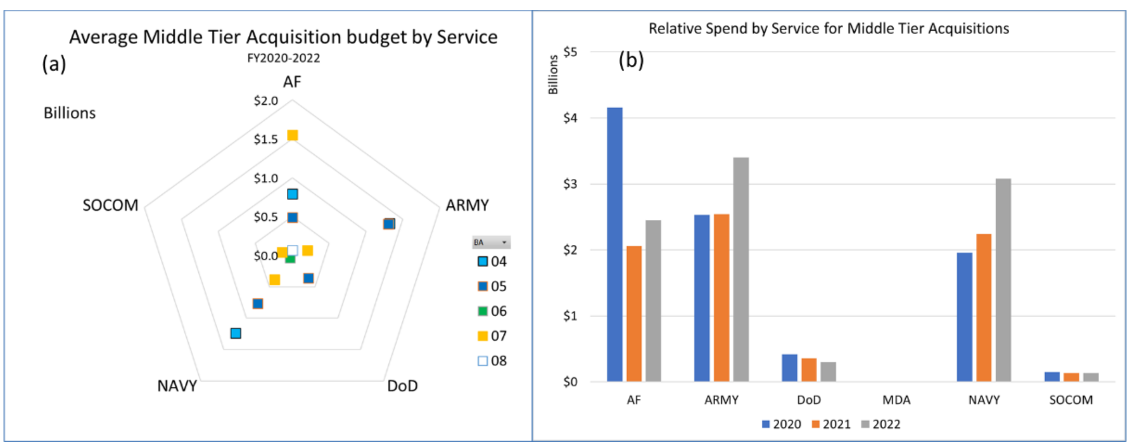


Figure 7. Resource allocation related to MTA projects.

Figure 7 shows large Air Force and Navy average investments, an initial investment surge by the Air Force, and increasing investments by the Army and Navy. The FY 2022 budgets show that MTA investment at the PE level is similar between the services. We specifically examined budget data at the MTA project level to differentiate between services.

---

[84] Mann-Whitney test, W-value = 1279, p-value = 0.267.

The results were that sum and average investments are statistically different[85] (α = 0.1) between services (Air Force, Army, or Navy), but not between BAs (BA 04, BA 05, BA 07).

Figure 8 shows MTA investments by commodity and type over start year at the MTA project level.

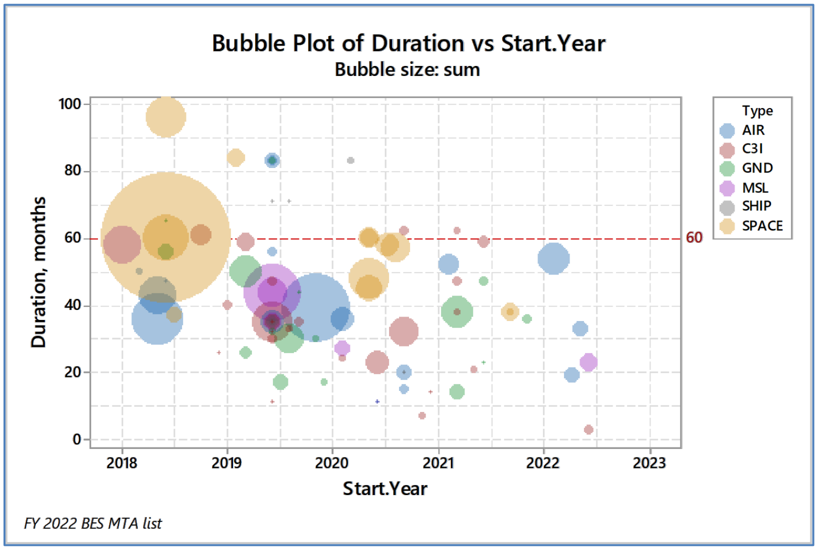


Figure 17. MTA Duration and Budget vs. Start Year

In Figure 8, a clear declining trend in large investments and longer durations is evident and confirmed by time series analysis. The conclusion is that the services are reducing project risk by focusing investments (smaller budgets and durations) and creating more programs to retire technical risks using rapid prototyping.[86] Figure 9 shows how schedules and budgets change by commodity type.

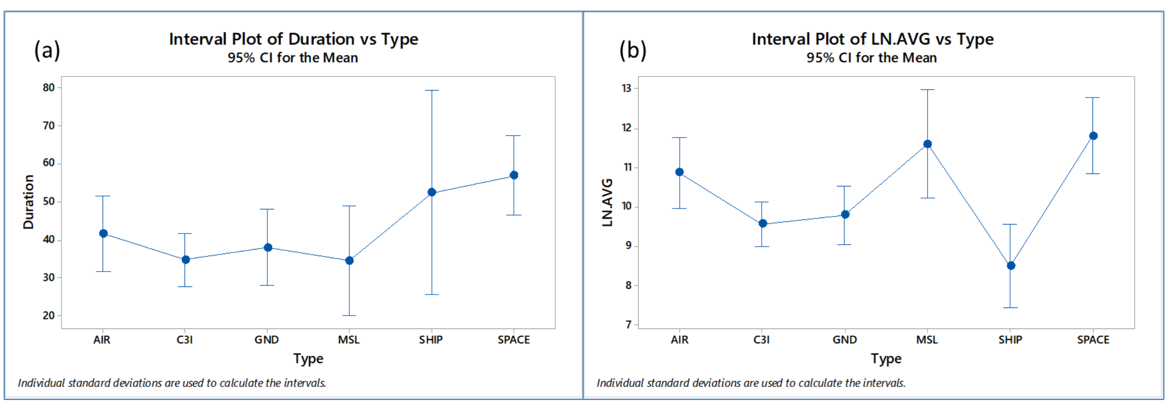


Figure 18. Schedule Duration and Budgets by Commodity Type

[85] Mood's Media test was used to compare medians; for sums and averages, Chi-square 0.72, p-value 0.083.
[86] Specifically, only four of 85 FY 2022 MTA projects were noted as Rapid Fielding MTAs.

Three MTA projects were excluded from budget analysis to meet ANOVA assumptions. MTAs are relatively indifferent to schedule; space commodities have the highest median durations, and ship-related MTAs have the largest variance. Average budgets are in Figure 9b and presented on a natural logarithm scale. Budgets show different groupings, with ship commodities having the smallest average budgets and C3I and ground commodity types being in a middle group.

Discussion

This DoD is evolving different approaches to MTAs. The Air Force was an early adopter, while the Navy was a later adopter of MTA project approaches, in part due to the different cultures and personalities noted by Riel (2020). These differences have reduced over time. Current MTA approaches generally have smaller budgets and shorter durations than earlier programs, reflecting lessons learned about the programmatic challenges associated with new acquisition approaches.

The services are employing MTA authorities to retire technical risks through rapid prototyping. A significant example of such use is the Air Force B-52 Commercial Engine Replacement Program, which is executing virtual prototype including different engine vendors and the prime integrator prior to attempting a physical prototype. A second example is the Army Integrated Visual Augmentation System, which has executed multiple physical prototypes with extensive soldier interaction at each prototype stage, resulting in rapid maturation of features and improved field reliability and performance. Both are novel prototyping approaches addressing different aspects of rapid capability development.

As previously noted, there is little research on MTAs. The FY 2022 dataset provides a detailed index for other researchers to explore MTAs and conduct detailed analyses, and for program offices to explore other creative and proven approaches to using MTAs to solve practical problems. The data used in this analysis was derived from public sources, and results and conclusions may differ if restricted or classified sources are used to replicate this work. Future research could include expanding research to include longitudinal studies of specific MTAs or MTA categories. The assessment of technical risk and system complexity affects the ability of program offices to properly scope MTA size, effort, and duration. Additional research is recommended to discover significant cost, schedule, and technical risk and complexity factors, which would be useful. Finally, research into changes in program office processes under MTA conditions would be useful to future program managers.

References

Cha, C. R., O Neill, S., Clement, M., Eyler, R., Fletcher, C., Irizarry, J., Kuhn, E., Panwar, M., & Sung, J. (2014). *Major automated information systems: Selected defense programs need to implement key acquisition practices* (GAO-14-309). GAO. https://apps.dtic.mil/dtic/tr/fulltext/u2/a598635.pdf

Chesbrough, H. W. (2003). *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business School Press. https://www.nmit.edu.my/wp-content/uploads/2017/10/Open-Innovation-the-New-Imperative-for-Creating-and-Profiting-from-Technology.pdf

Etemadi, A. H., & Kamp, J. (2021a). Market and contractor factors affecting rapid acquisition strategies. *Systems Engineering*. https://doi.org/10.1002/sys.21577

Etemadi, A. H., & Kamp, J. (2021b, May). Schedule risks associated with middle tier acquisition. *Proceedings of the 18th Annual Acquisition Research Program*. Acquisition Research Symposium, Monterey, CA. https://dair.nps.edu/handle/123456789/4397

FitzGerald, B., Lynn III, W., & O'Keefe, S. (2016). *Future foundry: A new strategic approach to military-technical advantage*. Center for a New American Security. https://www.cnas.org/publications/reports/future-foundry

Insinna, V. (2021, May 13). The F-22 will go away, eventually. But not before the Air Force gets comfortable with its successor. *Defense News*. https://www.defensenews.com/air/2021/05/13/the-f-22-is-going-away-eventually-but-not-before-the-air-force-gets-comfortable-with-its-successor/

Lord, E. M. (2019). *Operation of the middle tier of acquisition (MTA)*. DoD. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500080p.PDF?ver=2019-12-30-095246-043
Lord, E. M. (2020). *Operation of the adaptive acquisition framework*. DoD. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500002p.pdf?ver=2019-10-25-134150-283
Major Defense Acquisition Program Defined, Pub. L. No. 111–84, 10 U.S.C. § 2430 (2021). https://www.gpo.gov/fdsys/pkg/USCODE-2006-title10/html/USCODE-2006-title10-subtitleA-partIV-chap144-sec2430.htm
National Defense Authorization Act for Fiscal Year 2016, Pub. L. No. 114–92, S. 1356 (2015). https://www.congress.gov/114/plaws/publ92/PLAW-114publ92.pdf
Naval Postgraduate School. (2021, September). *Defense acquisition innovation repository*. Acquisition Research Program. https://nps.edu/web/acqnresearch/dair
Oakley, S. S. (2019). *DOD acquisition reform: Leadership attention needed to effectively implement changes to acquisition oversight* (GAO-19-439). GAO. https://www.gao.gov/assets/gao-19-439.pdf
Oakley, S. S. (2020). *Defense acquisitions annual assessment: Drive to deliver capabilities faster increases importance of program knowledge and consistent data for oversight* (GAO-20-439). GAO. https://www.gao.gov/assets/gao-20-439.pdf
Oakley, S. S. (2021). *Defense acquisitions annual assessment: Updated program oversight approach needed* (GAO-21-222). GAO. https://www.gao.gov/assets/gao-21-222.pdf
Riel, D. M. (2020). Faster acquisition: Putting the priority on speed. *Proceedings of the Seventeenth Annual Acquisition Research Symposium*, *I*, 27.
Stanley-Lockman, Z. (2021). From closed to open systems: How the US military services pursue innovation. *Journal of Strategic Studies*, *44*(4), 480–514. https://doi.org/10.1080/01402390.2021.1917393
Tate, D. M. (2016). *Acquisition cycle time: Defining the problem (revised)* (NS D-5762). Institute for Defense Analyses. https://www.jstor.org/stable/resrep22689?seq=1#metadata_info_tab_contents
Williams, T. (2005). Assessing and moving on from the dominant project management discourse in the light of project overruns. *IEEE Transactions on Engineering Management*, *52*(4), 497–508. https://doi.org/10.1109/TEM.2005.856572

Acquisition Warfare: A Proposal for a Unifying Concept

Lieutenant Commander Ryan Hilger—is an active duty Navy Engineering Duty Officer who served onboard USS Maine (SSBN 741; GOLD), as Chief Engineer of USS Springfield (SSN 761), and ashore with the CNO Strategic Studies Group XXXIII, at OPNAV N97, and on tours as an Engineering Duty Officer. He holds a BA in political science from the University of Kansas, an MS in mechanical engineering from the Naval Postgraduate School, and is a doctoral student in systems engineering at Colorado State University. His opinions are his own and do not reflect the official position of the Department of Defense. [rphilger@colostate.edu]

Abstract

The ongoing debate in the United States over defense acquisition reform highlights the complexity and evolution of the national security ecosystem. That complexity, explored using a first order system dynamic model, indicates that defense acquisition reform may be a so-called "super wicked" problem. Solutions to super wicked problems form a new class of solutions than traditionally found in the literature for defense acquisition reform. This paper asserts that defense acquisition reform is a super wicked problem and that adoption of an ecosystem model from the program office's perspective will yield new insights into ecosystem dynamics. Additionally, American adversaries, principally China and Russia, have used a variety of tactics and operations in systemic campaigns targeting the liminal space within the defense acquisition ecosystem. This paper proposes the unifying concept of acquisition warfare to better describe the set of adversary actions and how they disrupt the ability of program managers to successfully deliver their programs, not just systems, uncompromised within cost, schedule, and performance constraints.

Introduction

Over the last 70 years, the Department of Defense (DoD) acquisition system has continually evolved to meet perceived changes in the international threat environment, priorities from Congress of a new presidential administration, or the whims and preferences of key leaders. Researchers at the Center for Strategic and International Studies observed that within those 70 years, the DoD has initiated eight different acquisition reform cycles, split evenly between centralizing and decentralizing reforms (Dwyer et al., 2020). Today, the American defense establishment is again gripped by great power competition, simultaneously calling for faster action to retain American supremacy on the battlefield while bemoaning the lack of progress in acquiring weapons systems faster. The ebb and flow of changes mimics the patterns found in life, not the static, monolithic structure that we perceive the DoD to be.

The nature of these reforms and the cyclical patterns indicate that, despite professed desires, the national security establishment has not yet gotten defense acquisition reform "right." That psychological dissatisfaction with the status quo finds firmer theoretical grounding when viewed as a wicked problem, a term first coined by Horst Rittel and Melvin Webber in 1973 (Rittel & Webber, 1973). Levin et al. (2012) introduced a variation of wicked problems that especially fit the lingering discontent: super wicked problems. The particularly nasty planning problems exhibit additional characteristics that fit well with defense acquisition reform: urgency, lack of a single responsible entity to solve it, and humans acting as humans are wont to do—irrationally. As the problem evolves, so must the solution space, which means that we never solve the same problem twice.

To bring cohesion to the problem of defense acquisition reform and unify that problem with the latest round of international pressure, this paper offers two hypotheses. First, defense acquisition reform is a super wicked problem based on the behaviors and structure of the defense ecosystem. Second, a new theory of acquisition warfare represents a novel approach to understanding both the frustrations with reform and the avenues by which adversaries exploit the features of the ecosystem for their relative advantage. To support both hypotheses, this paper also develops a first attempt at a model of the defense ecosystem from the particular perspective of a defense acquisition program office.

Defense Acquisition Reform as a Super Wicked Problem

In their seminal 1973 paper on "Dilemmas in a General Theory of Planning," Horst Rittel and Melvin Webber defined a new class of planning problems as wicked problems, including an explicit mention of the new Planning, Programming, and Budgeting System developed under Secretary of Defense Robert McNamara (Rittel & Webber, 1973). At its core, the Defense Acquisition System (DAS), comprised of the Planning, Programming, Budgeting, and Execution (PPBE) system, the Joint Capabilities Integration and Development System (JCIDS), and the acquisition system, is a planning system. In its idealized form, planning is about explicit definition of a problem, articulation of desired goals or outputs, and the alignment of the fewest resources needed to accomplish the goal. In the DAS, executing the process of delivering a capability or system from start to finish is an idealized process that all in the defense acquisition and requirements space are familiar with. Rittel and Webber (1973), however, cast doubt on the efficacy of such systems: "And yet we know that such a planning system is unattainable, even as we seek more closely to approximate it. It is even questionable whether such a planning system is desirable." Indeed, those words remain equally true in 2022 as they were in 1973 as Congress launches a new commission to reform the PPBE process (Lineweaver et al., 2013; Serbu, 2021).

For these large scale planning, or wicked, problems, Rittel and Webber (1973) distilled 10 identifying characteristics: 1) no definitive formulation of the problem; 2) the problem does not stop, it just changes; 3) solutions are relatively good or bad; 4) there is no immediate test for efficacy of solutions; 5) every attempt at a solution is a "one-shot" operation since it changes the system; 6) the number of potential solutions are innumerable; 7) each problem is unique; 8) each wicked problem can be considered a symptom of another problem; 9) discrepancies have no single defining explanation; and 10) the planner has no right to be wrong (Rittel & Webber, 1973). In 2012, Levin et al. expanded Rittel and Webber's conceptualization of the wicked problem to encompass particular governance or policy planning problems where human behavior is irrationally biased toward short-term time horizons despite the more severe long-term impacts of those actions. These "super wicked" problems have four additional primary characteristics in addition to the original 10: "time is running out; those who cause the problem also seek to provide a solution; the central authority needed to address them is weak or non-existent; and irrational discounting occurs that pushes responses into the future" (Levin et al., 2012).

While Rittel and Webber (1973) originally characterized the Defense Acquisition System as a wicked problem, it fits better under the super wicked problem framework proposed by Levin et al. (2012). J. Ronald Fox et al.'s (2011) analysis in "Defense Acquisition Reform 1960-2009: An Elusive Goal" shows the overwhelming need for coordination amongst all stakeholders within the defense acquisition ecosystem, which reflects the lack of a centralized governance structure and the reaction to changes in the environment—mostly the Soviet Union—and the difficulty of producing an enduring solution (Fox et al., 2011).

Today, many of those same leaders still work within the defense acquisition ecosystem, and again the national security establishment espouses a driving need to reform the system in response to renewed global competition from a resurgent Russia and a rapidly growing China. The DoD's annual report to Congress for 2021, required by Congress since 2000, reports that the People's Republic of China has set a near-term military modernization goal of 2027 to provide additional, credible options for use against Taiwan as part of their longer-term goal of achieving a dominant military position by 2049 (*Military and Security Developments Involving the People's Republic of China*, 2021). The report certainly has a bias to it given the incentive for the DoD to inflate risks and consequences in an effort to secure additional funding from Congress, but the general content of the report can be independently confirmed by independent analysts on China and other open-source reporting. As a result, over the last few years, DoD officials, as summarized by the Congressional Research Service in their "Report to Congress on Great Power Competition," continue to stress that the time available to modernize the military is running out—China will surpass American military capabilities without significant investment and reform of all aspects of the defense ecosystem (*Renewed Great Power Competition: Implications for Defense-Issues for Congress*, 2022).

News articles from a one-week period in May 2021 alone highlight how different stakeholders have different perspectives on the issue, and example headlines range from "We Are Lost in the Woods on Defense Acquisition Reform" to "Acquisition Reform Is Making Rapid Progress, Defense Official Says" to "Just in: Pentagon 'Doubling Down' on Acquisition Reform" (Tadjdeh, 2021; Vergun, 2021; Welter, 2021). Other efforts from Congress over the last few years have required the DoD to examine reforms to the acquisition system writ large with the Section 809 panel, to smaller reform efforts for software acquisition practices, contracting options, and more (*Advisory Panel on Streamlining and Codifying Acquisition Regulations Volume 3 of 3 Summary of Recommendations*, 2019). The former Under Secretary of Defense for Acquisition and Sustainment, Ellen Lord, even enacted the most sweeping changes to Defense Systems Engineering in decades with the complete overhaul and reissuance of Department of Defense Directive 5000.01, taking the DoD from a traditional waterfall-centric systems engineering model to a "choose your own adventure" set of pathways for programs to choose from in an effort to streamline and accelerate acquisition of defense systems (*DoDD 5000.01 The Defense Acquisition System*, 2020).

Despite the clarion call to action for more aggressive reforms with respect to Chinese modernization, little progress seems to be made beyond small efforts with localized, and often temporal, results. Defense officials, Congress, presidential administrations, and the acquisition workforce all understand the pressing need to accelerate their efforts and deliver capability and capacity more rapidly, but the data shows that existing programs are continuing to execute their plans and Defense leaders have made little headway with Congress in divesting of legacy programs in favor of new technologies, changing acquisition strategies and pathways for major defense programs, and continue to have problems in meeting cost, schedule, and performance objectives ("High Risk Area - DOD Weapon Systems Acquisition," n.d.). In some cases, the DoD appears to be moving opposite from the direction of reform by slipping new programs further in the future, continuing to buy legacy systems as a result of delays to new programs, and sinking additional costs into these legacy programs from organizational inertia to modernize them to the limits of their available margins—the Concorde effect or sunk cost fallacy on a large scale (Arkes & Ayton, 1999). DoD weapon system acquisition first made the Government Accountability Office's (GAO's) High Risk list in 1990, providing a detailed longitudinal view of defense acquisition, but the GAO reports that despite strong leadership commitment in the last few years, progress remains significantly hindered ("High Risk Area - DOD Weapon Systems Acquisition," n.d.). Thus, despite strong Congressional support, committed DoD acquisition leadership and workforce, and a pressing, time-driven need, little change is occurring in the ecosystem—truly a super wicked problem. Perhaps a new approach to understanding the defense acquisition system is needed. Combining the disparate elements into a broader ecosystem model may yield new insights into how the defense acquisition ecosystem functions and reveal potential solution paths that could alter ecosystem behaviors and dynamics.

The Program Ecosystem

An ecosystem is a complex and coherent system of biophysical and social factors capable of adaptation and sustainability over time. While ecosystems generally conjure images of nature, the underlying principles govern human ecosystems as well. Ecosystems have structure which may or may not be directly observable (Margalef, 1963). The ecosystem may be observed indirectly through various metrics, behaviors, and trends within the ecosystem. In the case of defense acquisition, this includes measurable properties such as federal funding, employees, counts of weapons systems, etc., even though those metrics do not directly measure the structure and rules of the ecosystem. That structure, according to Ramon Margalef (1963), becomes "more complex, more rich, as time passes; structure is linked to history." The richness and evolution of the defense acquisition system, as chronicled by J. Ronald Fox et al. (2011), proves that, even though the system still seeks to produce the same outputs, the way in which it structures itself and alters the resource and information flows adapts and changes over time in response to various changes. Burch et al. (2017) developed the Human Ecosystem Model, shown in Figure 1, to better show general ecosystem dynamics and behaviors. Given that ecosystems are dynamic systems, the actors cannot produce observable behaviors or the measurable metrics without some exchange or flow of resources. Burch et al. (2017) identify six flows in the Human Ecosystem Model: individuals, energy, nutrients, materials, information, and capital (Burch et al., 2017). Other than nutrients, perhaps, all of these are applicable to driving action for producing decisions, capabilities, and outcomes in the defense acquisition system.

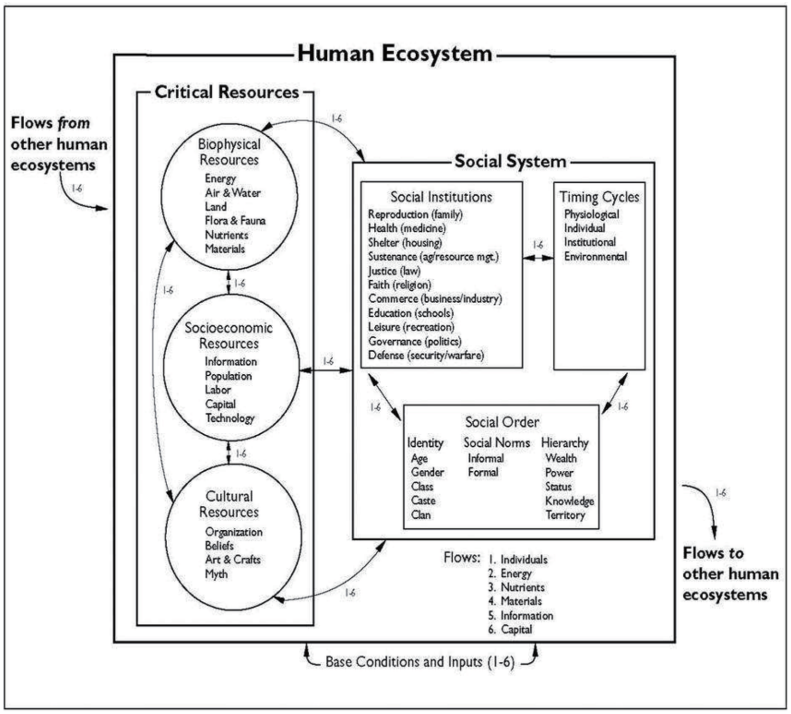


Figure 1. Human Ecosystem Model Developed by Burch, Machlis, and Force (2017)

In this context, *the program ecosystem is the coherent collection of people, processes, and systems working in the surrounding physical, cyber, and information domains to design, develop, produce, operate, and sustain national security systems and is viewed from the perspective of a defense acquisition program office*. That ecosystem can be modeled to show the dynamics that influence the behavior within that ecosystem. Figure 2, below, presents a simplified model of a defense program ecosystem. The simplified and local detailed models were developed primarily from the author's personal experience and required professional development activities as a generic, normative model of a defense acquisition program office, and do not necessarily represent the ecosystem of a specific program within the DoD.

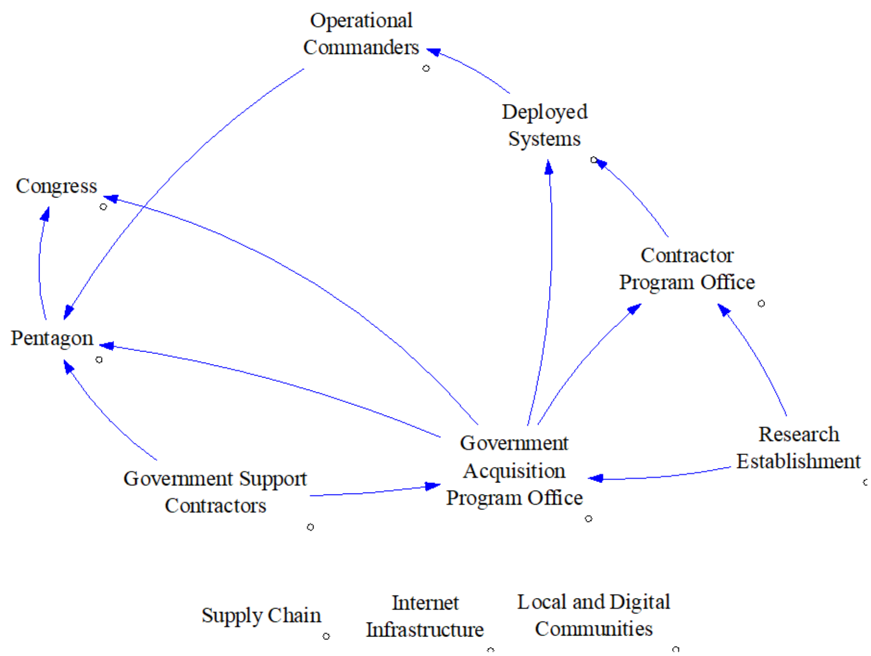


Figure 2. Simplified Program Ecosystem Model

The model reveals that a defense acquisition program does not operate in a vacuum, but rather within the broader environment, which in turn is impacted on the larger forces acting on the national security-industrial base, global supply chains, and the national and international environments. In turn, the entire ecosystem is supported by the physical and digital communities the workforces are a part of, the supply chains that provide individuals with their basic needs and the program with its material, and the internet that underpins the fabric of modern society. Thus, the program ecosystem is a complex adaptive system that exhibits emergent behaviors as the forces and flows of the ecosystem change over time. At the level of the program manager, the ability to control the cost, schedule, and performance of their given program is subject to the complexity of the ecosystem and the forces acting on the ecosystem at all levels, not just the deployed weapon system. Developing and generating effective combat power requires the flow of resources through the program ecosystem, notionally starting with requirements validation in the Pentagon, Congressional authorization and appropriation of funding, and expenditure of funds to design, develop, deploy, and sustain these systems. People at every node in the ecosystem have their own processes and procedures to execute to complete their step in the system value stream.

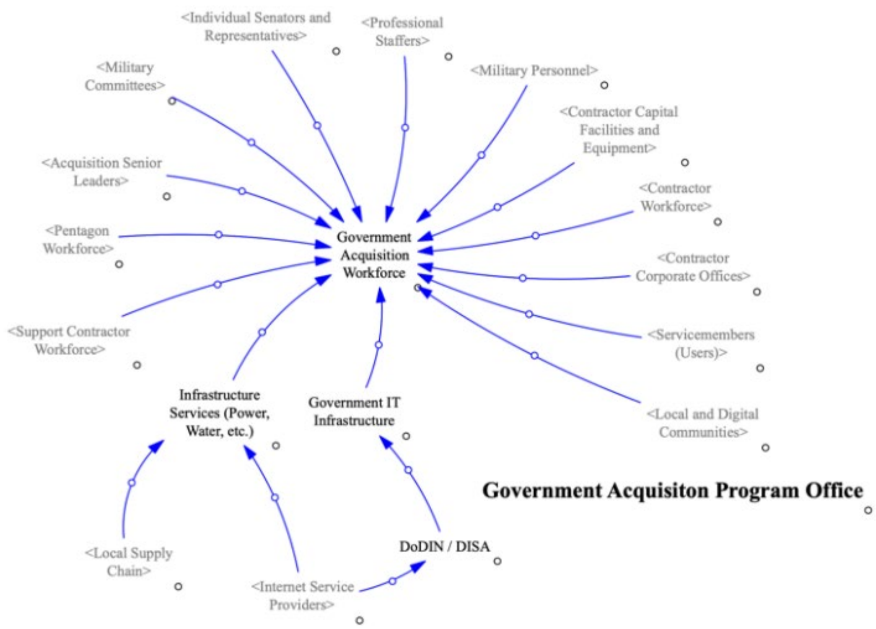


Figure 3. Detailed System Model of a Government Acquisition Program Office

The major actors in the ecosystem, the workforce of various organizations, do the daily work of defending the nation from conceiving of initial capabilities to designing and producing systems to conducting deployed operations. The relationships show how each of those major actors interact within the ecosystem. As the model shows, each of these people goes about their business dependent on the whole ecosystem. People have homes, are part of their local communities, are reliant on local infrastructure services and supply chains, and have digital lives depending on their local internet service providers. Thus, the forces acting on each of the people in a defense acquisition program ecosystem influence their daily behavior and their ability to focus on the program. Combined with the underlying infrastructure that supports human activity, the program ecosystem also represents a system view of a program's attack surface and the various propagation paths for vulnerabilities or other adversarial effects to disrupt the program or its deployed system. To provide better insight into how the ecosystem works, we provide a detailed model of several of the major actors and a brief description of the behaviors observed.

Government Program Office

Figure 3 shows the web of relationships that employees in a defense acquisition program must manage daily and the resources that go into supporting their daily work. Each day may find the program manager defending their program before Congress, meeting with leaders in the Pentagon to determine future years budget strategies and giving program updates, meeting with industry suppliers and their contractors, liaising with the servicemembers using their systems, and planning and scheduling modernization and deployment efforts with their commanders.

Contractor Program Office

Figure 4 shows the contractor's program office is equally complex, though the detailed model shows a different network of relationships that they must maintain and different forces that influence contractor behavior. As the contractor works the systems engineering lifecycle, they interact with suppliers at many levels, the government entities, including Congress through lobbying organizations, and the deployed systems, inclusive of the hardware, software, and users, to sustain the system.

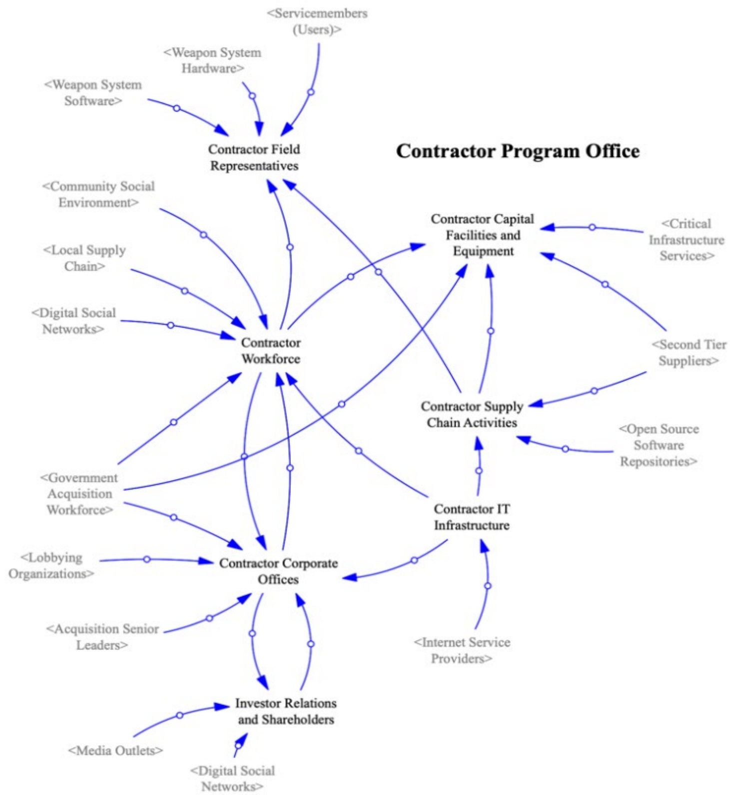


Figure 4. Detailed Model of a Contractor Program Office

Pentagon

The Pentagon, the long-standing nexus of defense acquisition funding and top-level requirements, operates in an equally complex web within the program ecosystem, as shown in Figure 5. Focusing specifically on the acquisition role in the Pentagon, the acquisition workforce and senior leaders work closely with Congress, the defense program offices, the business development offices of defense contractors, the media, and the military commanders to ensure that the defense program office has the resources to meet the requirements requested by those military commanders.

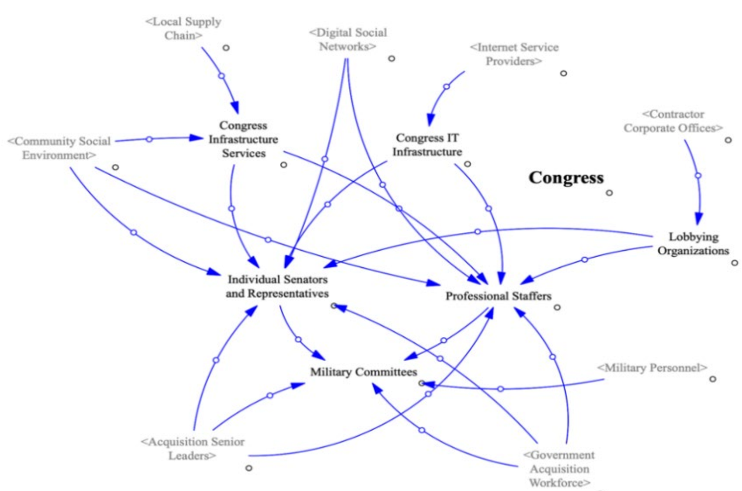


Figure 5. Detailed Model of the Pentagon Ecosystem

Congress

Congress retains the constitutional authority to authorize program, appropriate funds, and set broad policy guidance through changes to federal law. As the model shows, both the individual Senators and Representatives and their professional staffs craft the laws. Those laws authorize defense programs to exist and appropriate funds for the program to execute. This process is influenced by several external entities. The external entities come from the local community in Washington D.C., where most of the Congressional workforce resides, as well as from the acquisition workforce in the program offices and the Pentagon, and from various lobbying organizations for the defense industrial base.

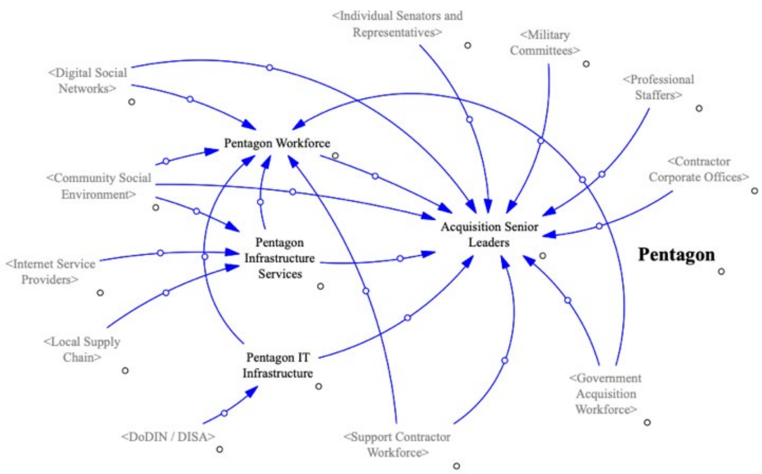


Figure 6. Detailed Model of the Congressional-Military Ecosystem

Deployed Weapon System

As shown in Figure 7, the deployed system is a scalable concept that ranges from an individual piece of equipment to a warship, satellite constellation, or long-range missile—anything produced through the Defense Acquisition System. In the program ecosystem, the deployed system can be defined as the hardware, software, the users of the system, and how they interact with the broader military environment. The servicemembers are influenced by the local community and its ability to sustain a population, their digital communities, and how they interact with the contractor's field representatives and the government program office to operate and sustain their systems.

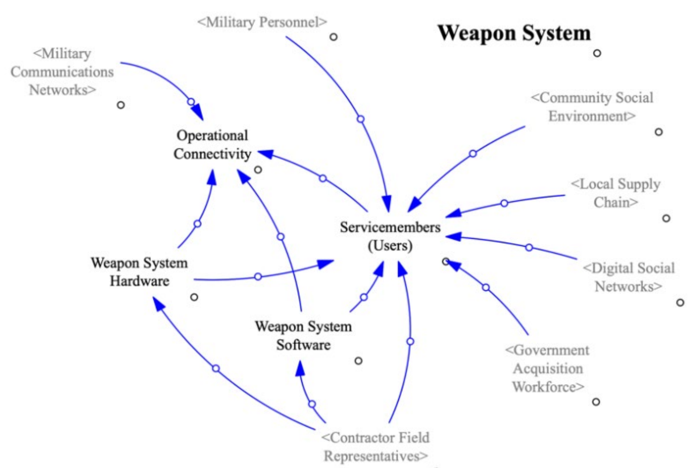


Figure 7. Detailed Model of a Deployed Weapon System

Research Ecosystem

The research ecosystem, modeled in Figure 8, represents the initial conception and early development of future capabilities. There are various actors, ranging from academia for basic research to mature research and development organizations for more complex prototypes and proofs of concept. Each of these organizational types have different incentives and influences to support and drive their research agendas. Like the other subcommunities within the program ecosystem, the researchers and their organizations are supported and influenced by their physical and digital communities.

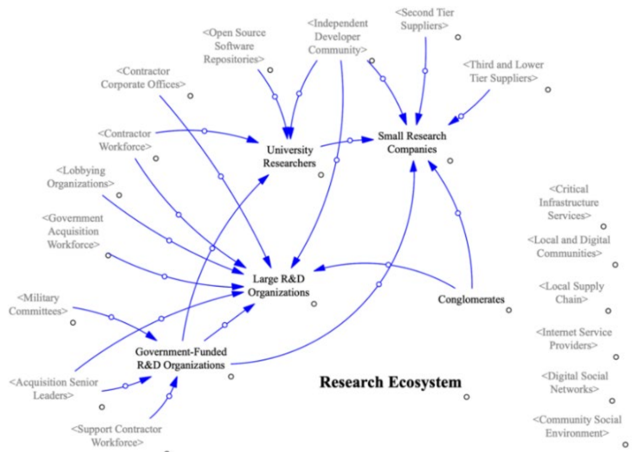


Figure 8. Detailed Model of the Defense Research Ecosystem

Local Community

Though often forgotten in the defense acquisition process, the local communities where the various program activities take place play a significant role in shaping individual and organizational behavior. The local community, shown in Figure 9, comprises two major parts: physical and digital. The physical communities are the neighborhoods and cities where people physically live, the civic and social activities we undertake during our lives, and the supply chains that provide for the basic needs and wants of the community. The digital communities, which include media outlets, have taken a major role in our lives and play an increasingly significant role in shaping our behaviors and attitudes, which spill into our work environment and shape the growing trend of digital nomadism (Reichenberger, 2017).

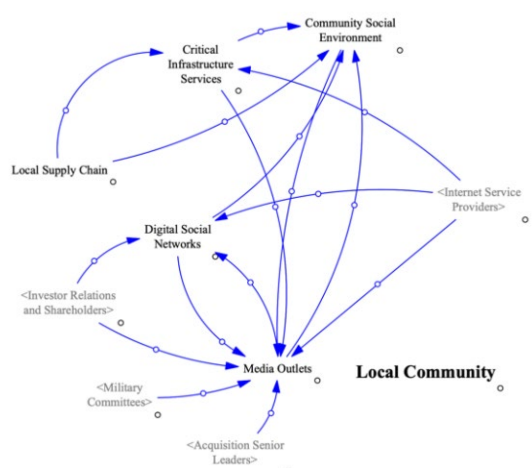


Figure 9. Detailed Model of the Local Community as It Supports the Program Ecosystem

In considering the hypothesis of defense acquisition reform as a super wicked problem, the complexities of the issues come through when viewed through the lens of the program ecosystem. Many recent reform proposals target aspects of the program ecosystem, but in light of the relationships between the entities, it shows how difficult implementing effective change can be. For example, the Section 809 Panel, chartered by the Fiscal Year 2016 National Defense Authorization Act, conducted a comprehensive study and produced 93 different recommendations for streamlining defense acquisition (*Advisory Panel on Streamlining and Codifying Acquisition Regulations Volume 3 of 3 Summary of Recommendations*, 2019). These recommendations primarily support the defense program office in the ecosystem model and do not consider the proposals and needs for reform of Congressional processes, JCIDS processes and requirements management in the Pentagon, improving the defense industrial base, etc., nor the previous recommendations of Congressionally chartered studies, the defense advisory boards, or industry studies.

The sheer volume of recommendations from the myriad studies is symptomatic of a deeper concern within the functioning of the program ecosystem. Rapport et al. (1985) state that the “signs or symptoms of distressed ecosystems do not generally appear in isolation” and that there are key indicators of overall ecosystem health that can be monitored. The authors cite several studies showing “reductions in species diversity, increases in nutrient leaching,” the “simplification of the structure of plant and animal communities … and loss of part or all of the inventory of nutrients,” or the “shift away from complex arrangements of specialized species toward the generalist … away from diversity in birds, plants and fish toward monotony” (Rapport et al., 1985). In the program ecosystem, we see the steady consolidation of the defense industrial base to a smaller number of large defense contractors (reduction in species diversity and simplification of the structure akin to the transition from polyculture to monocropping in agriculture) and an increasing share of defense dollars going toward those consolidated defense contractors (nutrient leaching or loss of nutrients; Berenson, 2021; *Department of Defense Report State of Competition within the Defense Industrial Base*, 2022; Jang et al., 2021). Rapport at al. (1985) continue, stating “there is an evident linkage among features of a distressed ecosystem… changes in primary productivity are linked with changes in nutrient availability” and that the symptoms of an ecosystem in distress can only be viewed in retrospect, akin to vital signs in medicine that indicate a disease has already advanced (Rapport et al., 1985). Adversaries seeking to slow the ability of the DoD to credibly and reliably develop and generate global combat power will necessarily target the program ecosystem at its weakest points or at the points where action may result in the greatest leverage or compounded effects. The concept of acquisition warfare provides a novel and analogous approach to address the complexity present in the ecosystem as recent approaches to climate change and global health have with the “One Health” and “One Medicine” initiatives that have undertaken to unify the disparate elements and theories within the ecosystem (Zinsstag et al., 2011).

Acquisition Warfare

The United States’ adversaries operate with the same forces in the global environment and recently have proven more adept at leveraging non-kinetic means at the liminal edge of conflict. Liminality represents the zone between detection and overt response and has rapidly grown to be the prime maneuver space for adversaries seeking to engage the United States at levels below those that would trigger escalatory actions, such as use of force, economic sanctions, etc. The United States, on the other hand, has clung to the traditional framework for the “range of military operations” as the defining framework for inter-state conflict, which leaves out the organizations, programs, and people in the United States that develop, deploy, and sustain our military forces for their use. Our adversaries have exploited this seam, the liminal zone, with great results. The figure below shows the zone at which most of our adversaries seek to operate.

Both China and Russia evolved their doctrines following the 1996 Taiwan Strait Crisis and collapse of the Soviet Union, respectively, to focus on the effective and efficient use of national resources to achieve national aims without provoking the United States into action. China's recent behaviors in the Western Pacific and the challenges they prevent for operating conventional military forces can be seen as the culmination of a successful campaign in the liminal zone to raise the capability and quantity of China's military forces to parity with the United States. Russia, while still weak, has exploited other liminal operations and tested under battlefield conditions in Estonia, Georgia, Syria, and Ukraine, in addition to other countries.

China, Russia, and other minor adversaries have shown repeated use of several tactics, which are found frequently as nations, corporations, organizations, and people maneuver in the liminal space. These tactics fall short of the traditional American view of the range of military operations but have significant impact on the capability and readiness of our forces, both now and in the future. These tactics can directly impact acquisition operations and place current and future programs and their systems at greater risk.

The tactics used generally run from clandestine operations to covert or ambiguous actions in the liminal space. Some tactics, if used too aggressively or overtly, may lead to immediate attribution and a proportional or escalatory response from the United States. Current, observable tactics in use by adversary nations include cyber warfare, industrial espionage and intellectual property theft, supply chain disruptions or compromise, lawfare, exploitation of humans, and information operations. Some of these tactics have been combined by adversaries to achieve specific objectives: rapid technological advancement and military modernization, sowing of distrust in program efficacy, and more. Acquisition warfare, as a concept, focuses primarily on the liminal actions and forgoes the impact of clandestine operations--the objective is to eventually provide program managers with a framework to actively defend against and counter adversary tactics that they can "see." Clandestine activities, while assessed to be ongoing, are primarily the domain of law enforcement and the intelligence community and generally only rise to the level of visibility for program managers and staff when the threat transitions to insiders, cyber-enabled access, etc. Thus, we define acquisition warfare as:

> Acquisition warfare is the set of tactics, operations, and campaigns to disrupt, delay, or deny an adversary effective research, development, production, or sustainment of current or future capabilities by means of clandestine and liminal actions designed not to elicit a response from the target nation.

Some may argue that acquisition warfare is simply Phase Zero operations by another name. However, the use of Phase Zero terminology conjures legacy definitions of military actions to shape the battlespace: peacetime deployments, presence operations, humanitarian assistance and disaster relief, community relations projects, etc. Phase Zero is focused on operations external to the United States. Joint Doctrine Note 1-19 introduced the competition continuum, which begins to describe parts of the framework introduced under acquisition warfare as "competition below armed conflict," though the broader scope focuses on the whole of nation—vice government—approach (*Competition Continuum [JDN 1-19]*, 2019). Acquisition warfare differs from traditional Phase Zero operations in that it necessarily includes a whole of government approach that is focused on defending future programs and systems, not influencing existing international relationships and operations today. Additionally, it aims to provide program managers with the necessary platform and agility to respond to changes in the acquisition environment.

Adversary Campaigns and Tactics

Acquisition warfare consists of several common tactics that adversaries bundle into short-term operations or longer-term campaigns to achieve a specific objective against a U.S. acquisition effort or to positively advance their own developmental programs. These tactics include cyber warfare, industrial espionage, supply chain disruption and compromise, lawfare, exploitation of people, and information operations.

China is the most prolific adversary in this space, and an ongoing analysis from the Center for Strategic and International Studies highlights the severity and widespread nature of these campaigns. The findings are worth quoting at length to show the complexity of the campaigns. The statistics represent 160 cases from 2000 to the present only in the United States and exclude more than 1,200 cases of intellectual property litigation against Chinese companies.

For those cases where we could identify actor and intent, we found:

- 42% of actors were Chinese military or government employees.
- 32% were private Chinese citizens.
- 26% were non-Chinese actors (usually U.S. persons recruited by Chinese officials).
- 34% of incidents sought to acquire military technology.
- 51% of incidents sought to acquire commercial technologies.
- 16% of incidents sought to acquire information on U.S. civilian agencies or politicians.
- 41% of incidents involved cyber espionage, usually by State-affiliated actors.
- This list is derived from open-source material and likely does not reflect the full number of incidents. Of the 160 incidents, we found that 24% occurred between 2000–2009 and 76% occurred between 2010–2021 (*Survey of Chinese Espionage in the United States Since 2000*, 2021).

In 2018, the Office of the Director of National Intelligence published a report on Foreign Economic Espionage in Cyberspace that described holistically the Chinese campaign and objectives for economic espionage: develop comprehensive national power, an innovation driven economic growth model, and rapid modernization of the military (*Foreign Economic Espionage in Cyberspace*, 2018).

Russian activities tend to be more focused on cyber warfare and information operations to generate effects at the national, vice defense, levels, such as the interference with the U.S. presidential elections in 2016 and the prolific use of botnets to further polarize the American electorate, with the resulting emergent effects rippling into the defense program ecosystem in the form of greater budget uncertainty, among others (Badawy et al., 2018). Though the two countries are unlikely to be cooperating closely in conducting acquisition warfare campaigns against the United States, the variability of adversary campaign tactics highlights the dynamic, multi-front nature of acquisition warfare.

Cyber Warfare

Cyber warfare is not strictly a military activity, as China's actions have shown and as Russia's use of cyber warfare in integrated campaigns has highlighted in recent years. From a U.S. national perspective, cyber warfare presents a direct threat to U.S. critical infrastructure that we rely on to support program ecosystems: the power grid, water and sanitation systems, health care, consumer good supply chains, and others (Genge et al., 2015) Many adversaries are active in this space, and the last few years have witnessed several non-state actor cyber campaigns against U.S. critical infrastructure and cleared defense contractors (*Chinese Gas Pipeline Intrusion Campaign, 2011 to 2013*, 2021; *Russian State-Sponsored Cyber Actors*

Target Cleared Defense Contractor Networks to Obtain Sensitive U.S. Defense Information and Technology, 2022; *Understanding and Mitigating Russian State-Sponsored Cyber Threats to U.S. Critical Infrastructure*, 2022). The Colonial Pipeline attack was the most recent and most public example of this, but other attacks from a drone on a power substation and poisoning of water systems have also happened in the last few years (Greenberg, 2021; Trevithick, 2021; Turton & Mehrotra, 2021).

From a program perspective, cyber warfare tactics and operations are key enablers to facilitate other objectives, whether the theft of intellectual property or plans, insertion of malicious code into weapon systems, or gaining of information to blackmail and compromise cleared personnel. The Cybersecurity & Infrastructure Security Agency, in conjunction with the National Security Agency and Federal Bureau of Investigation, provides detailed information on several nation-state actors conducting cyber warfare against the United States, in some cases including detailed descriptions of specific tactics and operating patterns.

In acquisition warfare, cyber operations may not be direct attacks against military weapon systems while deployed, but rather the deliberate compromise of program networks or supporting software (e.g., NotPetya, SolarWinds, Log4j, Microsoft Exchange, etc.) and hardware for the purposes of compromising the system before it is deployed, or to concurrently developed tailored countermeasures against our systems. While this largely falls under the new umbrella of program protection today, current program protection efforts do not fully account for the cyber-attack surface that critical program information is exposed to. In 2019, the Secretary of the Navy commissioned and released a Cybersecurity Readiness Review, which gives an accurate and likely little-changed picture of the current state of Navy defenses against this element of acquisition warfare (*Secretary of the Navy Cybersecurity Readiness Review*, 2019). Most likely, latent malware has been placed in every penetrated system that may be activated in the run-up or beginning of hostilities. This malicious infrastructure has likely already compromised much of the program ecosystems.

Industrial Espionage

Industrial espionage may be effected through cyberspace, physical access, human exploitation, or a combination of means. From a program ecosystem perspective, this manifests as the intrusion and exfiltration of controlled unclassified program and technical information, contractor proprietary information, and compromise of classified networks for the same. This may be done through cyber means such as compromising the network to gain access to sensitive information, insider threats to steal information, overt solicitation of key individuals with critical subject matter expertise through blackmail or job recruitment (the Thousand Talents program), and acquisition of specific corporations or their parent companies to gain access to sensitive information ("Committee on Foreign Investment in the United States Annual Report to Congress," 2020; Nakashima & Sonne, 2018).

Supply Chain Disruption

Over the last few decades, globalization and technological advances in information technology, manufacturing, and cost control have driven supply chains to be 1) increasingly global, 2) just in time, and 3) brittle in the face of disruption. As the COVID-19 pandemic proved, lack of inventory is only one factor that can cause disruptions to supply chains. In the case of the national security ecosystem, supply chains are those that furnish the systems in development and deployment, the items necessary for the program to complete their mission, including the business and IT environments, and the local supply chains that ensure that the workforce has their basic needs met, both personally and for their family, and thus can contribute and focus fully in the work environment. Supply chain challenges have grown to the point that in February 2021, President Biden signed Executive Order 14017 to establish the

Supply Chain Disruption Task Force to address six national critical supply chains (*Executive Order on America's Supply Chains*, 2021)

Program managers cannot control the local supply chains, but they do have significant stake in the operation, resiliency, and efficacy of the supply chains that enable program action and the systems they deploy. Frequently, defense program offices and their contractor program office counterparts have little visibility into the program's overall supply chain below the first tier or two of subcontractors (Nothacker, 2021). Program managers must understand how adversary actions against supply chains impact the cost, schedule, and performance of the programs under management. In recent years, China, especially, has forcefully disrupted supply chains through campaigns to control certain sectors of the market, such as rare earth metals, compromise manufacturing supply chains to enable future access through cyber means to gain access to the program ecosystem (Dreyer, n.d.; Robertson & Riley, 2021). Given the extensive Soviet infiltration of the United States during the Cold War, the program ecosystem should be considered compromised already and subject to exploitation at the adversary's desire (Zhuk, 2022).

Lawfare

First defined by retired Air Force General Charles Dunlap in 2008, lawfare is "the strategy of using--or misusing--law as a substitute for traditional military means to achieve a warfighting objective." While normally exploited by non-state actors, non-governmental organizations, and others specifically to address human rights violations and similar issues, the tactics have come into increasing use to achieve various effects. China, naturally, has a widespread lawfare campaign to secure territory in the South China Sea (the Nine Dash Line and anchoring it to international law is a classic example of lawfare). Russia, as well, has used lawfare successfully to hold off the United Nations and other international bodies to allow it to act unencumbered in Crimea, Georgia, Estonia, and other states. At the corporate level, the exploitation of people for the Thousand Talents programs can bring the full weight of Chinese financial and legal resources to bear to tie a company up in expensive litigation over trade secrets, intellectual property, etc.

From an acquisition warfare perspective, lawfare targets individuals and corporations. Within each program ecosystem, there are certain individuals, whether key leadership or subject matter experts, who, if forced to leave the program, could significantly affect the ability of the program to develop, deliver, or sustain capabilities. For example, if there are a few key scientists who understand hypersonics and are critical to the ongoing development programs, their departure could place the program at undue risk. Through other means, China or other adversaries may have accumulated sufficient information, such as from information operations, cyber warfare (i.e., the Office of Personnel Management data breach), and others, to create and proffer false claims and charges against individuals, thus embroiling them in legal battles and significantly degrading their productivity and leadership within the program ecosystem. Russia recently demonstrated a similar use of lawfare in the lawsuits against several computer security researchers who had worked to expose the connections of Alfa Bank, and Russian Bank, to President Donald Trump's campaign organizations (Devlin, 2021). Such use of lawsuits, similar to the rising prevalence of doxxing in American culture today, may have a chilling effect on the ability to research and attribute cyber-attacks or other actions in the future (Calabro, 2018).

Human Exploitation

China, again, is the most prolific adversary exploiting American workers for gain. All intelligence services continue clandestine operations to use insiders, recruited agents, and others to steal secrets (*The China Threat — FBI*, n.d.). China goes further. Much further. Originally starting with the Thousand Talents Program to bring overseas knowledge and talent into the Chinese domain, the efforts expand across nearly all sectors, ministries, and organizations and at the national, regional or provincial, and corporate levels to comprise 43 programs at the national level and more than 200 at lower levels (*CSET Chinese Talent Program Tracker*, n.d.).

These are rarely clandestine activities. As Nicholas Eftimiades (2020) reports, most of the human exploitation activities are conducted openly, with minimal to no espionage tradecraft, and with overtly stated objectivesClick or tap here to enter text.. This is the recruitment of people from academia, corporations, and government to provide information or services to China, share research or plans with China first, or simply move to China to work in a company, laboratory, or university there for higher pay. Many are Chinese nationals operating without cover names or stories and are conducting business openly, but one in four are American citizens recruited by Chinese officials. Defense programs and especially their supporting contractors are key targets given the industries we are involved in and the advanced technologies we work with.

Cultivation of a potential source often begins with a combination of information operations through social media and cyber activities to gain a better understanding of who the best targets for recruitment may be. Personnel advertising an active security clearance and program affiliation on LinkedIn may become targets for increased information gathering to allow for an eventual approach and recruitment, as happened to a former CIA officer (*Clearance Holders Targeted on Social Media*, n.d.).

Information Operations

All adversaries conduct information operations for the purpose of enhancing their own international stature or capabilities or degrading U.S. capabilities. This normally manifests in news articles, botnets to shape algorithmic search and filters, and campaigns to publicize new capabilities, but can also include the lower-level actions of pressuring U.S. companies to comply with Chinese laws if they serve or have business in China—think of YouTube, Facebook, and Google, for example, having to adjust software to prevent access to certain sites, as captured by Peter Singer and Emerson Brooking (2019) in *LikeWar*Click or tap here to enter text..

In the acquisition warfare space, this primarily comes out as creating immense pressure on the American acquisition system and its people through a sense of being behind Chinese, Russian, or other adversary capabilities in a particular area. The announcements of test results for hypersonic anti-ship cruise mission tests, fractional orbital bombardment systems, or earlier operational capability deployments of unmanned systems, etc. have allowed our adversaries to use other elements of the national security-industrial complex and American media to put pressure on acquisition programs to deliver, which compounds with high-profile test failures or other program setbacks as the individuals in the program ecosystem seek to reap more benefits or resources than the ecosystem can reasonably or sustainably provide at that time (Hitchens, 2021; Newdick & Rogoway, 2021; Pollack, 2022; *Russia Test-Fires New Hypersonic Tsirkon Missiles from Frigate, Submarine*, 2021).

Conclusions and Areas for Future Research

This paper sought to unify the ongoing frustrations with defense acquisition reform and accelerating great power competition by, first, framing defense acquisition reform as a super wicked problem and, second, proposing the novel framework of the program ecosystem and acquisition warfare to provide a new lens from which to shape future actions at all levels of the defense acquisition ecosystem. The program ecosystem model provides a first attempt at describing the systemic forces that drive the behaviors of the program ecosystem. That understanding may yield new insights into how proposed changes will drive ecosystem behaviors and identify more optimal points for effecting change to achieve a desired outcome. As Levin et al. (2012) identify for global climate change, understanding a problem with super wicked characteristics will help policymakers better identify the solution sets that will be both palatable to other actors in the ecosystem and achievable within the current operating characteristics of that ecosystem.

Acquisition warfare and the program ecosystem, as a new concept, offer a multitude of avenues for future research. For program managers, adapting acquisition warfare and developing a program-specific ecosystem model, preferably through ethnographic and other forms of research to develop a functional casual loop diagram or dynamical model, will allow for a better understanding of the program's overall attack surface and help identify the limits of the program manager's influence over various behaviors in the ecosystem. Further research is needed to better identify and document adversary campaigns and their impacts on programs and tracing out the effects of those campaigns through the program ecosystem model to highlight the dynamism of them. The growing analytical capabilities in the field of network science will provide insight into network dynamics and help inform potential changes to the sociotechnical design of the ecosystem. The possibilities for future research in this new field are extensive.

References

Advisory Panel on Streamlining and Codifying Acquisition Regulations Volume 3 of 3 Summary of Recommendations. (2019).

Arkes, H. R., & Ayton, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin*, *125*(5), 591. https://doi.org/10.1037/0033-2909.125.5.591

Badawy, A., Ferrara, E., & Lerman, K. (2018). Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign. *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2018)*, 258–265. https://doi.org/10.1109/ASONAM.2018.8508646

Berenson, D. (2021). *The evolving geography of the U.S. defense industrial base - War on the rocks*. War on the Rocks. https://warontherocks.com/2021/09/the-evolving-geography-of-the-u-s-defense-industrial-base/

Burch, W. R., Machlis, G. E., & Force, J. E. (2017). *The structure and dynamics of human ecosystems: Toward a model for understanding and actions*. Yale University Press.

Calabro, S. (2018). From the message board to the front door: Addressing the offline consequences of race- and gender-based doxxing and swatting. *Suffolk University Law Review*, *51*. https://heinonline.org/HOL/Page?handle=hein.journals/sufflr51&id=65&div=&collection=

Chinese Gas Pipeline Intrusion Campaign, 2011 to 2013. (2021, July 21). CISA. https://www.cisa.gov/uscert/ncas/alerts/aa21-201a

Clearance Holders Targeted on Social Media. (n.d.). Retrieved March 29, 2022, from https://www.fbi.gov/investigate/counterintelligence/the-china-threat/clearance-holders-targeted-on-social-media-nevernight-connection

Committee on Foreign Investment in the United States Annual Report to Congress. (2020). In *Committee on Foreign Investment in the United States*. https://home.treasury.gov/system/files/206/CFIUS-Public-Annual-Report-CY-2020.pdf

Competition Continuum (JDN 1-19). (2019). https://www.jcs.mil/Portals/36/Documents/Doctrine/jdn_jg/jdn1_19.pdf

CSET Chinese Talent Program Tracker. (n.d.). Retrieved March 29, 2022, from https://chinatalenttracker.cset.tech/

Department of Defense Report State of Competition within the Defense Industrial Base. (2022). https://media.defense.gov/2022/Feb/15/2002939087/-1/-1/1/STATE-OF-COMPETITION-WITHIN-THE-DEFENSE-INDUSTRIAL-BASE.PDF

Devlin, B. (2021, November 18). Alfa Bank lawsuits, Sussman indictment reignite battles over Trump organization computer links.. *The Washington Post*. https://www.washingtonpost.com/national-security/alfa-bank-trump-russia-sussman/2021/11/18/b10e498c-4726-11ec-b8d9-232f4afe4d9b_story.html

DoDD 5000.01 The Defense Acquisition System. (2020). DoD. https://www.esd.whs.mil/DD/.

Dreyer, J. T. (n.d.). *China's monopoly on rare earth elements—and why we should care*. Foreign Policy Research Institute. Retrieved March 29, 2022, from https://www.fpri.org/article/2020/10/chinas-monopoly-on-rare-earth-elements-and-why-we-should-care/

Dwyer, M., Tidwell, B., & Blivas, A. (2020). *Cycle times and cycles of acquisition reform.* Center for Strategic and International Studies. https://www.csis.org/analysis/cycle-times-and-cycles-acquisition-reform

Eftimiades, N. (2020). *Chinese espionage: Operations and tactics*. Vitruvian Press.

Executive Order on America's Supply Chains. (2021, February 24). https://www.whitehouse.gov/briefing-room/presidential-actions/2021/02/24/executive-order-on-americas-supply-chains/

Foreign Economic Espionage in Cyberspace. (2018). https://www.dni.gov/files/NCSC/documents/news/20180724-economic-espionage-pub.pdf

Fox, J. R., Allen, D. G., Lassman, T. C., Moody, W. S., & Shiman, P. L. (2011). Defense acquisition reform 1960-2009: An elusive goal. In *Defense Acquisition Reform*. Center of Military History, United States Army.

Genge, B., Kiss, I., & Haller, P. (2015). A system dynamics approach for assessing the impact of cyber attacks on critical infrastructures. *International Journal of Critical Infrastructure Protection*, *10*, 3–17. https://doi.org/10.1016/J.IJCIP.2015.04.001

Greenberg, A. (2021). *A hacker tried to poison a Florida city's water supply, officials say.* Wired. https://www.wired.com/story/oldsmar-florida-water-utility-hack/

High Risk Area - DOD Weapon Systems Acquisition. (n.d.). In *GAO High Risk List*. GAO. Retrieved March 26, 2022, from https://www.gao.gov/highrisk/dod-weapon-systems-acquisition

Hitchens, T. (2021, November 29). *It's a FOBS, Space Force's Saltzman confirms amid Chinese weapons test confusion*. Breaking Defense. https://breakingdefense.com/2021/11/its-a-fobs-space-forces-saltzman-confirms-amid-chinese-weapons-test-confusion/

Jang, W. J., Sanders, G., & Holderness, A. (2021, December 2). *2021 defense acquisition trends: Topline DoD trends after a half decade of growth.* Center for Strategic and International Studies. https://www.csis.org/analysis/2021-defense-acquisition-trends-topline-dod-trends-after-half-decade-growth

Levin, K., Cashore, B., Bernstein, S., & Auld, G. (2012). Overcoming the tragedy of super wicked problems: Constraining our future selves to ameliorate global climate change. *Policy Sciences*, *45*(2), 123–152. https://doi.org/10.1007/S11077-012-9151-0

Lineweaver, C. H., Davies, P. C. W., & Ruse, M. (2013). What is complexity? Is it increasing? In *Complexity and the Arrow of Time*. Cambridge University Press.

Margalef, R. (1963). American naturalist on certain unifying principles in ecology. *The American Naturalist*, *97*(897), 357–374.

Military and Security Developments Involving the People's Republic of China. (2021).

Nakashima, E., & Sonne, P. (2018). China hacked a Navy contractor and secured a trove of highly sensitive data on submarine warfare. *The Washington Post*. https://www.washingtonpost.com/world/national-security/china-hacked-a-navy-contractor-and-secured-a-trove-of-highly-sensitive-data-on-submarine-warfare/2018/06/08/6cc396fa-68e6-11e8-bea7-c8eb28bc52b1_story.html

Newdick, T., & Rogoway, T. (2021, December 14). *Russia unveils stealthier version of its S-70 "Hunter" unmanned combat air vehicle*. The Warzone. https://www.thedrive.com/the-war-zone/43504/russia-unveils-stealthier-version-of-its-flying-wing-unmanned-combat-air-vehicle

Nothacker, D. (2021). Supply chain visibility and exception management. In *Disrupting Logistics* (pp. 51–62). Springer. https://doi.org/10.1007/978-3-030-61093-7_5

Pollack, J. (2022, January 3). *Why do US hypersonic missile tests keep failing? They're going too fast*. Defense One. https://www.defenseone.com/ideas/2022/01/why-do-us-hypersonic-missile-tests-keep-failing-theyre-going-too-fast/360276/

Rapport, D. J., Regier, H. A., & Hutchinson, T. C. (1985). Ecosystem behavior under stress. *The American Naturalist*, *125*(5), 617–640. https://www.jstor.org/stable/2461475

Reichenberger, I. (2017). Digital nomads – A quest for holistic freedom in work and leisure. *Annals of Leisure Research*, *21*(3), 364–380. https://doi.org/10.1080/11745398.2017.1358098

Renewed Great Power Competition: Implications for Defense-Issues for Congress (Vol. 92). (2022). https://crsreports.congress.gov

Rittel, H., & Webber, M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*, 155–169. https://www-jstor-org.ezproxy2.library.colostate.edu/stable/4531523?seq=1#metadata_info_tab_contents
Robertson, J., & Riley, M. (2021, February 12). *Supermicro hack: How China exploited a U.S. tech supplier over years*. Bloomberg. https://www.bloomberg.com/features/2021-supermicro/
Russia test-fires new hypersonic Tsirkon missiles from frigate, submarine. (2021, December 31). Reuters. https://www.reuters.com/business/aerospace-defense/russia-test-fires-new-hypersonic-tsirkon-missiles-frigate-submarine-2021-12-31/
Russian State-Sponsored Cyber Actors Target Cleared Defense Contractor Networks to Obtain Sensitive U.S. Defense Information and Technology. (2022, February 16). CISA. https://www.cisa.gov/uscert/ncas/alerts/aa22-047a
Secretary of the Navy Cybersecurity Readiness Review. (2019). https://www.wsj.com/public/resources/documents/CyberSecurityReview_03-2019.pdf?mod=article_inline
Serbu, J. (2021). *Pentagon's ponderous budget process is next target for Congressional reform.* Federal News Network. https://federalnewsnetwork.com/defense-main/2021/12/pentagons-ponderous-budget-process-is-next-target-for-congressional-reform/
Singer, P. W., & Brooking, E. T. (2019). *LikeWar: The weaponization of social media*. Mariner Books.
Survey of Chinese Espionage in the United States Since 2000. (2021). Center for Strategic and International Studies. https://www.csis.org/programs/technology-policy-program/survey-chinese-linked-espionage-united-states-2000
Tadjdeh, Y. (2021, May 12). Just in: Pentagon 'doubling down' on acquisition reform. *National Defense Magazine*. https://www.nationaldefensemagazine.org/articles/2021/5/12/just-in-pentagon-doubling-down-on-acquisition-reform
The China Threat — FBI. (n.d.). Retrieved March 29, 2022, from https://www.fbi.gov/investigate/counterintelligence/the-china-threat
Trevithick, J. (2021). *Likely drone attack on U.S. power grid revealed in new intelligence report (Updated)*. The Drive: The Warzone. https://www.thedrive.com/the-war-zone/43015/likely-drone-attack-on-u-s-power-grid-revealed-in-new-intelligence-report
Turton, W., & Mehrotra, K. (2021, June 4). *Colonial pipeline cyber attack: Hackers used compromised password*. Bloomberg. https://www.bloomberg.com/news/articles/2021-06-04/hackers-breached-colonial-pipeline-using-compromised-password
Understanding and Mitigating Russian State-Sponsored Cyber Threats to U.S. Critical Infrastructure. (2022, January 11). CISA. https://www.cisa.gov/uscert/ncas/alerts/aa22-011a
Vergun, D. (2021, May 14). *Acquisition reform is making rapid progress, defense official says*. DoD. https://www.defense.gov/News/News-Stories/Article/Article/2607639/acquisition-reform-is-making-rapid-progress-defense-official-says/
Welter, T. (2021, May 6). We are lost in the woods on acquisition reform. *Defense News*. https://www.defensenews.com/opinion/commentary/2021/05/06/we-are-lost-in-the-woods-on-acquisition-reform/
Zhuk, S. I. (2022). *KGB operations against the USA and Canada in Soviet Ukraine, 1953-1991*. Routledge.
Zinsstag, J., Schelling, E., Waltner-Toews, D., & Tanner, M. (2011). From "one medicine" to "one health" and systemic approaches to health and well-being. *Preventive Veterinary Medicine*, *101*(3–4), 148–156. https://doi.org/10.1016/J.PREVETMED.2010.07.003

Estimating Middle Tier Acquisition Schedule Risk

Dr. Amir Etemadi—Amir Etemadi received a Doctor of Philosophy degree in electrical engineering from the University of Toronto, Toronto, Canada. He is an Assistant Professor in the School of Engineering and Applied Science in the Department of Electrical and Computer Engineering and the Department of Engineering Management and Systems Engineering, and teaches masters- and doctorate-level courses in electrical engineering and engineering management. He is the Principal Investigator on several current research projects and has supervised over 50 masters and doctoral candidates for degrees in both electrical engineering and engineering management. [etemadi@gwu.edu]

Dr. John Kamp—John Kamp received a Doctor of Engineering degree in engineering management from the George Washington University. Kamp teaches masters-level courses in project and program management and is a faculty Graduate Research Advisor in the School of Engineering and Applied Sciences. He is a retired Naval Submarine Officer with extensive experience in research and development and program management. His research interests include engineering management, maritime systems, and acquisition system research. Kamp is a Fellow in the Royal Institution of Naval Architects and a member of several professional associations. [jckamp2018@gwu.edu]

Abstract

Congress recently created Middle Tier Acquisition (MTA) programs, which provide the military services rapid prototyping and fielding pathways with new program flexibilities and an explicit schedule constraint. The services are executing multiple MTAs, resulting in a set of MTA experiments related to development, execution, and governance. There is little published information on MTA performance; we use public data to quantify planned schedules. We introduce a quantified schedule risk measure based on Monte Carlo simulations. The simulations provide insights into MTA programs' schedule risk and program performance relative to a statistically based reference.

Research Results Statement: This research provides quantitative assessments of the effectiveness of Middle Tier Acquisition policy on schedule growth.

Keywords: Middle Tier Acquisition, defense acquisition, innovation

Introduction

Five economic and strategic policy changes occurred over the last 50 years affecting defense acquisitions. First, most research and development today is performed outside the United States. In 1960, the United States funded nearly 70% of world research and development (Sargent, 2018). By 2019, the percentages were reversed, and the U.S. defense share had shrunk to 3% of total global spend (Sargent & Gallo, 2021). Second, technical innovation shifted to primarily commercial sponsorship, and the Department of Defense (DoD) is now competing for emergent technologies (Sargent & Gallo, 2021). Third, the supporting industrial base of defense-unique suppliers shrank, affecting defense market competition and innovation (Etemadi & Kamp, 2021). Fourth, U.S. military strategy to emphasize technological and operational superiority (Grant, 2016), predisposing the United States to technology-dependent military operations. Finally, U.S. military operational priorities evolved to include military operations against peer competitors and non-state actors, and non-combat missions on a global scale. Gansler and Lucyshyn stated in 2010 that the "DoD's normal way of doing business … is totally incompatible with adversaries using available commercial technologies in new and different ways."

In 2016, Congress enacted new laws, referred to as Middle Tier Acquisitions (MTAs), addressing institutional (cultural) barriers, providing expedited processes available to speed execution, increasing discretionary authorities, and imposing direct accountability to deliver prototypes or fielded systems within five years of approval (National Defense Authorization Act, 2015).

Background

Reforms over the last 50 years tended to focus on authorities, governance, and performance (Fox, 2011) instead of the organizational culture. Williams (2005) considered that poor defense program performance resulted from systemic failures, in particular when conventional program management approaches were used for *complex, uncertain, and time-constrained* programs. Bower and Hout (1988) considered companies as systems, and shorter schedule durations (cycle times) came from improving organizational processes, such as adopting flexible manufacturing and focusing on small production runs responsive to customer demands.[87]

In 1992, the General Accounting Office[88] (GAO) observed that important programs struggled with cost and schedule growth and technical performance challenges, while successful programs avoided the "oversell and resulting performance bias" (GAO, 1992) of typical programs. Flyvbjerg (2006) noted most program forecasts are biased by psychological ("optimism bias") and political ("strategic misrepresentation") interests and proposed reference class forecasting to improve forecasting performance. Grau et al. (2017) found forecasting performance associated with organizational factors, such as early response to unfavorable trends and incentives for program management and control. Klein Woolthius et al. (2005) identified cultural issues such as regulatory and normative failures and groupthink or lock-in–type failures as limiting innovation policies and responses. Weber and Rohracher (2012) identified systemic failure causes for slow fielding of development projects, including early lock-in to suboptimal technology,[89] and challenges with change and adaption.[90]

Prior research identified factors related to "fast-to-field" programs, such as an urgency of need, senior leader sponsorship, and rapid access to available funding (Van Atta et al., 2016), and program strategy decisions associated with shorter schedules include using proven systems or developing and fielding systems with incremental performance improvements (Tate, 2016). The overall competence or proficiency of an organization[91] affects their ability to plan and execute development and production (Jaifer et al., 2020).

Technical maturity is commonly defined in the DoD using an ordinal scale of Technology Readiness Levels (TRLs; Mankins, 2009). In this context, maturity is a marker of successful use in conditions approaching the intended environment and purpose, where increasing TRL values indicate increasing maturity. Technical maturity is a commonly identified cause of schedule growth, where schedule growth decreases as a product becomes more mature (Katz et al., 2015). Markham and Lee (2013) analyzed commercial new product development; as seen with the DoD, radical innovations take longer and are more likely to experience schedule and cost growth relative to incremental innovations.[92] Dougherty (2018) identified technological maturity and sponsorship[93] as key factors for rapid prototyping, but additionally noted urgency of need

[87] These remain appropriate for improving process efficiency.
[88] Now the Government Accountability Office.
[89] Identified as a policy coordination failure.
[90] Identified as a reflexive failure.
[91] Organization in this context includes suppliers and "testers."
[92] Radical innovations are presumed to be less technically mature than incremental innovations.
[93] Sponsorship includes championing and protecting or obtaining resources.

and a compelling demonstration as important for rapid fielding. He recognized that rapid programs need only be "good-enough" to satisfy the sponsor's immediate performance needs.

New product development literature discusses gated management methods, such as Agile Stage Gate (Cooper, 2017). Lingens et al. (2016) proposed using estimates of potential impact and uncertainty to frame management decisions. Van Oorschot et al. (2018) argued that how fast a new product goes to market is a trade-off between cycle time and product quality, and requires understanding how many new (unexpected) tasks are identified in the front-end stage, how many are discovered just prior to the decision gate, and how many customers will wait for the new product. This approach makes sense when trying to maximize profit. The recent pandemic provides some insights into *rapid* new product development. Battaglia et al. (2021) identified three key factors for rapid development: *technical competence*, a*gile work practices* in order to produce and demonstrate a working prototype, and a*ccess to networks*[94] and scale to meet production, certification, and commercialization demands. This is consistent with the findings of Hsiao et al. (2017) on better new product performance when firms exploit core capabilities (competencies) and first-mover strategies.

Jaifer et al. (2020) developed an aerospace new product development case study, grouping schedule-related factors into *complexity* and *proficiency* categories, with uncertainty a subset of complexity. Bearden et al. (2012) developed a system complexity index for space systems[95] that showed a strong association between cost growth and the complexity index. Jaifer et al. (2020) decomposed complexity into technological complexity,[96] organizational complexity,[97] and environmental complexity.[98] The acquisition process itself exhibits inherent complexity. Wirthlin (2009) developed a discrete event simulation model of the DoD acquisition process reflecting systemic complexity by including process interdependencies as programs competed for finite resources.[99] In large scale systems-of-systems, the interdependencies may manifest as interoperability or integration issues, with discovery and correction needing exquisite engineering discipline (Garrett et al., 2011). Prescriptions include increasing system resiliency (Roberts et al., 2016), increasing inherent reliability, and reducing system complexity by increasing system commonality, modularity, and use of standards (Jovel & Jain, 2009).

A significant source of program uncertainty is the DoD preference for technical superiority, and the system technical maturity reflected in the *technical debt* that the program must retire (Boehm & Behnamghader, 2019). Patil and Bhaduri (2020) argued for incremental requirements, constrained ("frugal") developments, and frequent interaction with users ("fast") to reduce development uncertainty and improve product innovation, Williams's (2005) program uncertainty is related to technical uncertainty and may be addressed by reusing existing technologies (Eiband et al., 2013). Peters et al. (2017) showed that new technology development challenges reduce technology maturity, increasing program uncertainty. Program schedule uncertainty is also related to program complexity, so organizational factors such as contract types and intentional schedule overlaps matter (Jaifer et al., 2020).

---

[94] Battaglia et al. (2021) included external funding as part of the network.
[95] Named the Complexity-Based Risk Assessment (CoBRA).
[96] Technological complexity includes factors such as technical maturity, number of requirements, components, and functions.
[97] Organizational complexity includes factors such as contract types, program phase overlaps, team size, and number of technical disciplines.
[98] Environmental complexity includes factors such as legal constraints, numbers of stakeholders and suppliers, and competition levels.
[99] In Wirthlin's (2009) model, most programs never progressed to production.

In prior research, we developed regressions showing that Major Defense Acquisition Program[100] (MDAP) schedules were related to the research and development budget size, whether the program depended on another MDAP, the reuse of existing or commercial technology, the type of software development, and whether or not the program is joint with another service (Etemadi & Kamp, 2021b). There are other policy-related decisions associated with schedules and schedule growth such as percent research and development budget remaining at program start (Jimenez et al., 2016), using incremental development (Mortlock, 2019), and contract type selection (General Services Administration, 2019).

Schedule risk definitions differ in the literature, ranging from the likelihood to achieve a predicted duration (Dubos et al., 2007) to an estimate of likelihood and consequence (Tao et al., 2017). Browning (1998) used causal loop representations to identify likely sources and consequences of schedule delays, and showed how uncertainty drives risk. Earned-value methods are used to estimate schedule performance risk within a known project scope and budget (Swartz, 2008). Such simulations require detailed work project schedules and duration uncertainty distributions as inputs.

The existing literature describes organizational factors related to and qualitative attributes of rapid new product development. The military services, led by the Department of the Air Force, are discovering how to rapidly deliver new capabilities to the field to employ these new authorities. This paper is based on research we conducted for the Acquisition Research Program under Grant No. 12936478. Our research focused on rapid acquisitions, including MTAs, their program structures and products, and what acquisition strategy decisions were made to achieve schedule performance. In this paper we discuss use of simulations to understand schedule-related issues associated with constrained duration programs.

Material and Methods

While some MTA programs are delivering products now, most are not yet reporting sufficient data for analysis. Given the relative newness of MTAs, we used publicly reported event dates to characterize schedule variances for MTA and MDAPs,[101] and developed Monte Carlo simulations seeded with these recent program data. The simulations help develop insights into MTA performance relative to MDAPs.

We used publicly available data sources including GAO annual weapon system assessments (Oakley, 2020) and released Selected Acquisition Reports (SARs).[102] We started with the 2020 GAO annual weapon system assessment (Oakley, 2020; n = 63), eliminating entries with insufficient data (n = 3) and programs changing structures (n = 2). As each commodity type has unique development issues, we further reduced this to consider only air and missile commodity types,[103] leaving 27 entries, shown in Table 1.

Table 1. Selected GAO 2020 Air and Missile Programs

Program ID	Service (SVC)	Commodity Type (Type)	Program ID	Service (SVC)

[100] These are traditional large defense acquisition programs. They are now called Major Capability Acquisitions.
[101] Formerly Major Defense Acquisition Programs, now called Major Capability Acquisitions (Lord, 2020).
[102] See Washington Headquarters Services (2022).
[103] A commodity type represents the product, in this case an aircraft system or a missile system.

					Commodity Type (Type)
APT	AF	AIR	ITEP	Army	AIR
B2DMSM	AF	AIR	VC25.RECAP	AF	AIR
AARGM-ER	Navy	MSL	VH92	Navy	AIR
CIRCM	Army	AIR	JAGM	Army	MSL
CRH	AF	AIR	B52RMP	AF	AIR
F15EPAWSS	AF	AIR	IFPC.Inc2	Army	MSL
CH-53K	Navy	AIR	PrSM	Army	MSL
KC46A	AF	AIR	P8A.INC3	Navy	AIR
IRST.BLK2	Navy	AIR	ARRW*	AF	MSL
SDB.INC2	AF	MSL	B52CERP*	AF	AIR
UH-1N.REP	AF	AIR	F22CP*	AF	AIR
MQ25	Navy	AIR	HCSW*	AF	MSL
MQ4C	Navy	AIR	F35	DOD	AIR
NGJ-MB	Navy	AIR	** Indicates MTA*		

The DoD has specific acquisition pathways (Lord, 2020). We defined a generic schedule consisting of a program start, a decision gate to start development,[104] a critical design review (CDR), a decision gate to start production,[105] a declaration of initial operational capability (IOC), and intervals between these events, as shown in Figure 1.

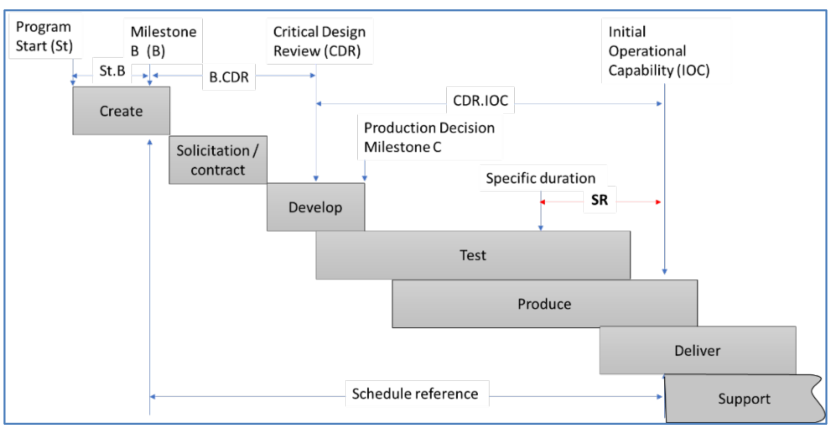


Figure 1. Notional Schedule and Schedule Reference

[104] Called Milestone B in the Major Capability Acquisition pathway, it is also development contract award, formally the start of development.
[105] Called Milestone C in the Major Capability Acquisition pathway, it is marked by award of an initial or low-rate production contract.

In Figure 1, the interval labels identify the starting and ending events, such as St.B being the interval in months between program start and the development start decisions. Table 2 summarizes the intervals used in this paper.

Table 2. Interval Definitions

Interval	Description
St.B	Interval between initiation (St) and development decision (start of Engineering and Manufacturing Development phase, or Milestone B)
B.CDR	Interval between development decision (Milestone B) and Critical Design Review (CDR)
CDR.IOC	Interval between Critical Design Review (CDR) and Initial Operational Capability (IOC)
B.IOC	Interval between development decision (Milestone B) and Initial Operational Capability (IOC)

MTA programs are structured as either a Rapid Prototyping or Rapid Fielding program, as shown in Figure 2.

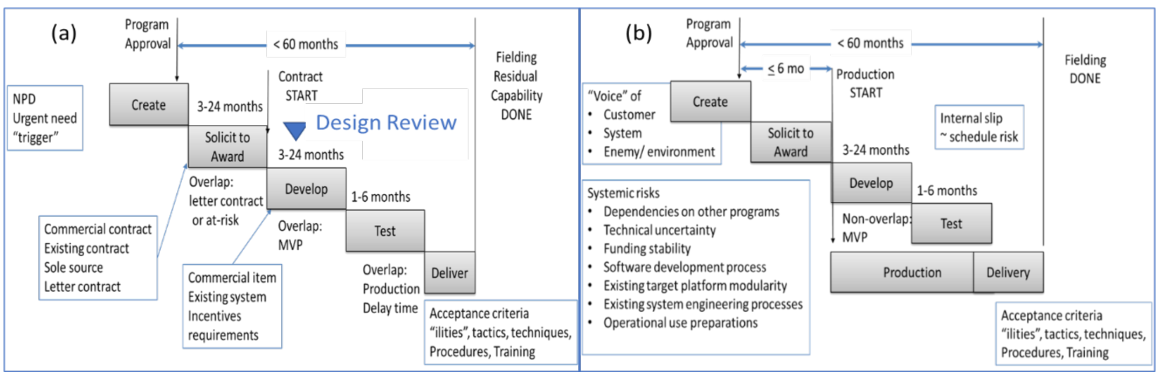


Figure 2. MTA Schedule Models[106]

While the overall sequencing is like Figure 2, there is an explicit schedule constraint of 60 months between program start and completion forcing early decisions to achieve this constraint as shown in the text blocks.[107] The Rapid Prototyping MTA (Figure 2a) is the most common model. The product is a prototype or residual capability. In this model, program approval is the equivalent of development start (Milestone B), and fielding is IOC. There is no explicit CDR requirement; typically, a design review occurs early in development. Unlike the Rapid Prototyping model, the Rapid Fielding MTA (Figure 2b) is intended for delivery of operational products and forces an early product start within 6 months of program approval.

The program *schedule* is the interval between development start (Milestone B) and IOC.[108] We define the program schedule as

[106] NPD = New Product Development, MVP = Minimum Viable Product.
[107] Schedule speed results from use of commercial-type contracts or Other Transaction Agreements, and adaptation of commercial or near-commercial products.
[108] This interval is also known as the program cycle time (the variable names Cycle.Mo in the data set).

$$schedule = B.CDR + CDR.IOC \quad (1)$$

Schedule will generally equal the interval between Milestone B and IOC (B.IOC), unless a program has no reported CDR. We do not include the time prior to the development decision (St.B) in the overall schedule as this is prior to a formal development commitment, and includes planning and precontract activities. We modeled schedule as both a sum and as an interval (B.IOC), and compared results. Additionally, we did not decompose the interval after CDR to IOC (CDR.IOC) further as real programs differ in both sequences and events during this interval, with parallel[109] execution of various development, testing, production, and deployment activities. We defined *schedule risk* (SR in Figure 2) as a measure of the remaining schedule to IOC, equivalently the likelihood of exceeding a specified schedule duration:

$$schedule\ risk\ (SR) = 1 - p(schedule\ \leq specific\ duration) \quad (2)$$

The right-hand-side probability is the cumulative distribution function for the overall or reference schedule.

We performed Monte Carlo simulations for each interval and for the overall schedule for the Table 1 subset, simulating normal and Weibull distributions for both MDAP and MTA program types. The MTA schedule risk simulation depends on two assumptions: the model distribution and the likelihood of exceeding the schedule constraint. For the first assumption, if MTAs are *equally likely* to experience schedule contraction or growth, then assuming schedules are normally distributed is reasonable. In practice, few programs deliver early, and many programs complete later than initially planned, so skewed distributions[110] were used to simulate reference schedule and interval distributions. The second assumption is the likelihood of MTA schedule duration exceeding 60 months. In this case, we assumed that acquisition executives (decision authorities) may restructure struggling programs, so we did not explicitly restrict MTA schedule durations from exceeding 60 months. Instead, we used schedule variance estimates from the GAO data to seed Monte Carlo simulation models created in Excel. Figure 3 shows an example set of normal cumulative distribution functions for MTA programs where the expected (planned) durations vary from 60 months (schedule.MTA) to 24 months (24.MTA) with an overlaid set of hypothetical intervals.

---

[109] Often called concurrency; a similar concept is fast-tracking.

[110] We used Weibull distributions (McCool, 2012) to simulate skewed distributions. Minitab 18 was used to calculate scale, shape, and threshold parameters from empirical data fits.

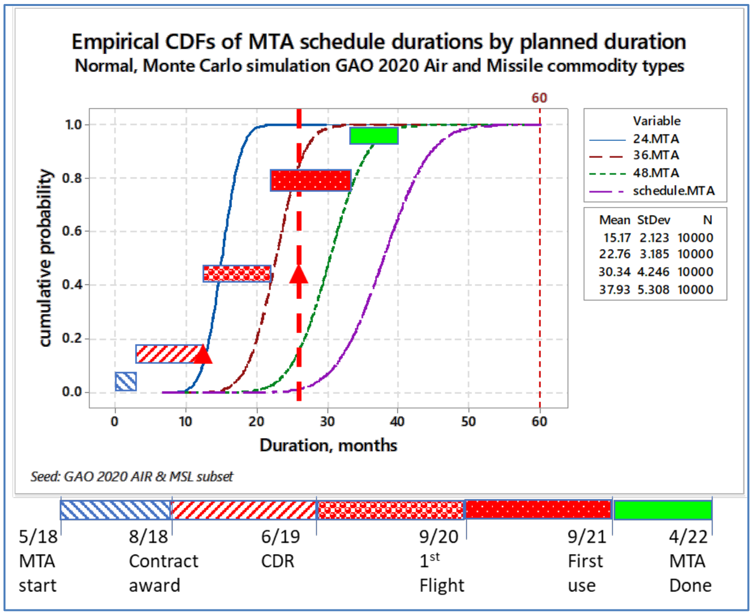


Figure 3. Example Mapping of Key Events to MTA Simulation Model

In Figure 3, intervals are *head-to-tail*, where the end of the final event corresponds to the overall planned duration. The red triangles are when an event occurred relative to program start (duration, months) *at the cumulative probability of the interval*. We see that the latest event occurred later than planned and can visually estimate how late a program might be without corrective actions. The green bar is a margin set by the program office, making the example MTA duration 40 months. The vertical dotted line is the duration at which first flight occurred (10 months late), and the red triangle on the red dotted line indicates the actual event completion date.

In Figure 3, first use corresponds to IOC and occurs at about 34 months, showing a planned schedule of nearly 36 months (36.MTA). In this example, the schedule risk is 0.6, meaning that the program has a 0.6 chance of *not* making the 40-month schedule without corrective action, but is likely to complete in the next 2 years (42–48 months total duration, or 2–8 months late), close to the original plan plus margin. There is little chance of the schedule exceeding 60 months, unless it crosses the 60-month (schedule.MTA) curve.

Results and Discussion

We sorted the Table 1 programs by MDAP and MTA and calculated descriptive statistics sorted by acquisition type (MDAP or MTA). Table 3 summarizes data set interval statistics.

Table 3. Interval Simulation Descriptive Statistics

Interval	Type	Mean	StDev	Scale	Shape	Threshold
St.B	MDAP	28.48	22.46	0.6683	24.04	-0.2923
	MTA	2.00	1.826	20.14	28.19	-25.62
B.CDR	MDAP	22.41	17.28	1.252	22.81	1.2
	MTA	17.67	4.51	51.59	162.4	-144.3
CDR.IOC	MDAP	83.33	43.88	1.91	52.34	34.17
	MTA	23.00	8.54	59.63	347.9	-3243
B.IOC	MDAP	102.50	56.3	1.495	89.37	21.67
	MTA	39.25	10.14	1962	13909	-13865
Cycle.Mo	MDAP	110.22	45.12	1.616	79.08	39.14
	MTA	25.25	18.98	2.093	36.12	-6.494

The distribution statistics reflect the small number of MTA programs (4) in the data set. We tested these intervals for goodness of fit against both normal and Weibull distributions, and were acceptable at a significance of 0.01. Figure 4 shows the histograms for these intervals.

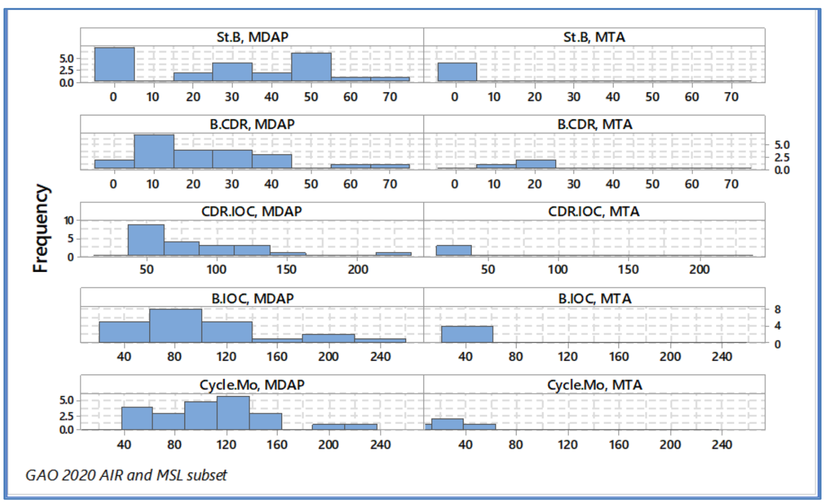


Figure 4. Empirical Interval Histograms

We developed Monte Carlo simulations of interval durations in Excel using interval data from the Table 1 programs. We ran 10,000 normal and Weibull simulations for each Table 3 interval and present selected Weibull simulation results in Figure 5.

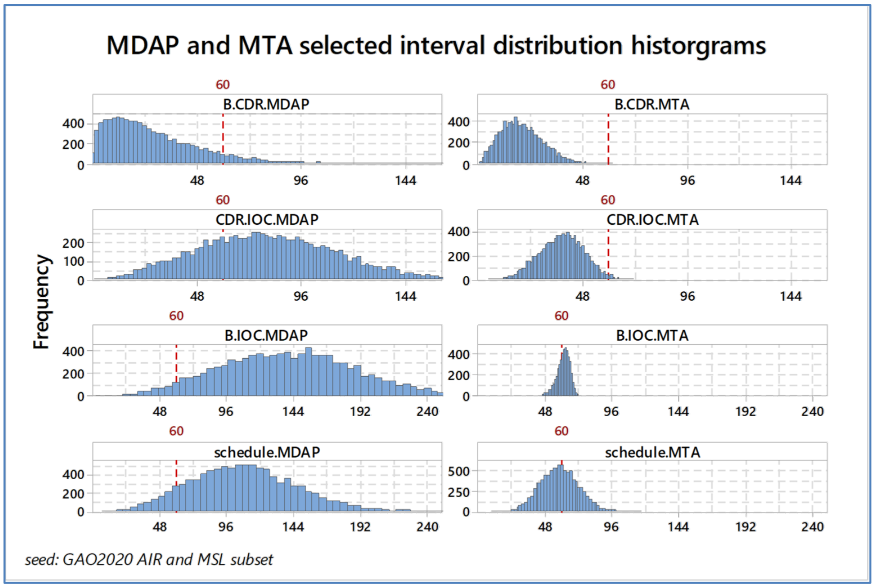


Figure 5. Weibull MDAP and MTA Interval Simulation Results

The Figure 5 simulations show the compression of MTA intervals and schedules relative to traditional MDAP programs. The right-skew of development and design (B.CDR) is more pronounced than that of the production to delivery (CDR.IOC) phase, meaning MDAP and MTA schedule durations are largely due to CDR.IOC. In all simulations, the MTA variance is much smaller than traditional MDAP programs. Figure 6 shows cumulative distribution functions for selected intervals and schedules for MDAPs and MTAs.

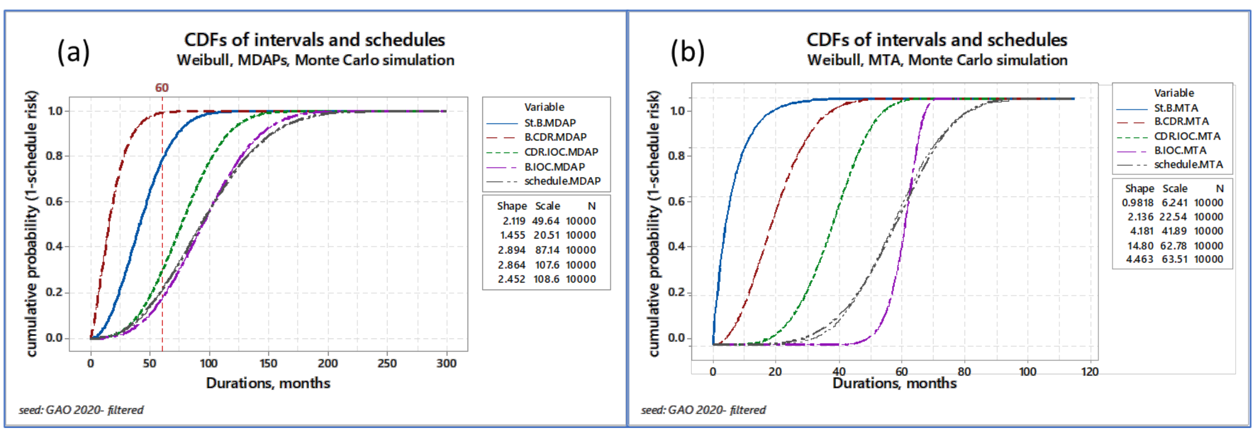


Figure 6. Cumulative Distribution Function Simulation Results

The time from program start to development start (St.B) is quite fast for MTAs, meaning that rapid contracting and award approaches are essential for an MTA.[111] Figure 6a shows that

[111] In principle, average MTA might be finished before the average MDAP is under contract.

the B.CDR is similar in shape to the interval between program start and Milestone B (St.B, blue curve). In Figure 6b, the MTA B.CDR curve is quite steep compared to Figure 6a, meaning MTA programs should prefer technologies closer to actual use (less technical uncertainty) than normally used by a MDAP.

The steepness of the CDR.IOC curve for MTA (red curve) emphasizes the inability of an MTA to accommodate system complexity within duration constraints. The program must complete integration and demonstration in about 2 years, while a MDAP will take about four times as long to proceed from CDR to initial fielding. This may reflect the different testing and initial production issues related with MDAPs. The MTA CDR.IOC curve is also steeper, meaning the system must be much less complex to produce and field than a MDAP system. We compare MDAP and MTA schedule performance in Figure 7.

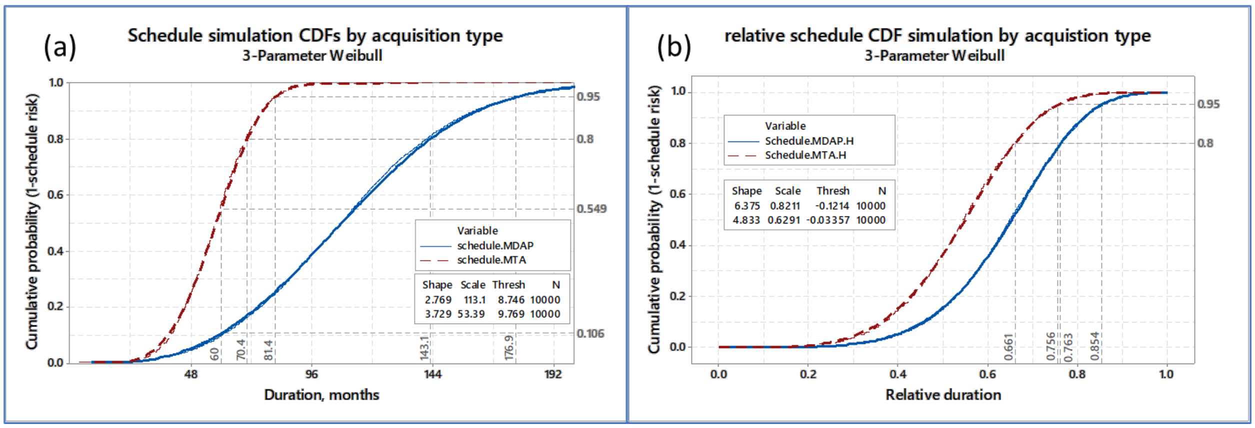


Figure 7. MDAP and MTA Schedule Duration Simulations

By inspection, Figure 7a shows MTAs have less schedule risk than MDAPs at the same *absolute* schedule duration. In Figure 7b, both schedule curves are scaled by the maximum duration for each distribution, resulting in duration fractions between 0 and 1. Figure 7a shows neither MDAPs nor MTAs are likely to take less than 2 years to complete. The MTA model was allowed to exceed 60 months due to random variation. The simulation suggests that MTAs are likely to exceed 60 months by less than 10 months without additional controls, but unlikely to exceed 84 months. Additionally, both Figures 7a and 7b show that MTAs have lower absolute (7a) or relative (7b) schedule risk than MDAPs. We conclude that the MTA schedule risk is less than an MDAP.

Conclusions

The simulation results show that DoD MTAs are unlikely to complete in less than 2 years, but can deliver within the 60-month limit. Program offices can adjust planning and execution to meet an explicit schedule constraint. The key program attributes that help achieve these results are short times to contract award (development start), and intentionally reducing program complexity and technical uncertainty. The short time to contract award can include both using commercial-type contracting methods and reducing contractual requirements. The MTA product design should not require extensive technical risk reduction and should complete in less than 12–18 months (B.CDR). The interval after design complete (CDR.IOC) has the most effect on schedule duration, meaning that program offices should simplify production, test, and certification gates, and plan additional schedule margin during this phase. The simulations show that adding schedule margin *at the end* of a program is prudent.

The simulations provide a way to assess program performance without requiring in-depth understanding of program plans. The ability to compare an overall duration and plan against an expected distribution and schedule risk allows discussions about why a schedule is faster or slower than expectations.

The development and application of schedule risk simulations to assess likely MTA performance is an original contribution of the paper. The research highlights the importance of reducing program complexity and uncertainty to reduce program schedules. The indirect quantification of program complexity and uncertainty in terms of program schedule allow discussions and trades to reduce the schedule risk associated with a particular program plan.

This research is applicable to DoD air and missile commodity type MDAP and MTA acquisition programs. The results are based on publicly released data. Different conclusions may result if these methods are applied to more complete data sets, other commodity types, or to programs with unstable requirements and resources. Future research opportunities include applying these methods to restricted data sets and comparing internal program assessments to these refined models, validation with future programs, and association of department, program office, and commodity type cultural factors to MTA schedule performance.

Credit Author Statement

Amir Etemadi: Funding acquisition, Supervision, Project administration, Writing - Review & Editing. John Kamp: Funding acquisition, Conceptualization, Methodology, Investigation, Data Curation, Formal analysis, Validation, Visualization, Writing - Original Draft.

Funding sources

This material is based upon work supported by the Acquisition Research Program under Grant No. 12936478. The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Department of Defense nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. government.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the colleagues who provided us feedback and our families for allowing us to take advantage of this opportunity. We thank the George Washington University for supporting our research efforts and helping us through proposal development and grant execution. We also thank the Acquisition Research Program and the Naval Postgraduate School for supporting and encouraging our research.

References

Battaglia, D., Paolucci, E., & Ughetto, E. (2021). The fast response of academic spinoffs to unexpected societal and economic challenges: Lessons from the COVID-19 pandemic crisis. *R&D Management*, *51*(2), 169–182. https://doi.org/10.1111/radm.12444

Bearden, D., Cowdin, M., & Yoshida, J. (2012). Evolution of complexity and cost for Planetary Missions throughout the development lifecycle. *2012 IEEE Aerospace Conference*, 1–12. https://doi.org/10.1109/AERO.2012.6187411

Boehm, B., & Behnamghader, P. (2019). Anticipatory development processes for reducing total ownership costs and schedules. *Systems Engineering*, *22*(5), 401–410. https://doi.org/10.1002/sys.21490
Bower, J. L., & Hout, T. M. (1988). Fast-cycle capability for competitive power. *Harvard Business Review*, *66*, 110.
Browning, T. R. (1998). 2.2.2 Sources of schedule risk in complex system development. *INCOSE International Symposium*, *8*, 213–220. https://doi.org/10.1002/j.2334-5837.1998.tb00031.x
Bruchey, W. J. (2012). *A comparison of earned value and earned schedule duration forecast methods on Department of Defense major defense acquisition programs* (Join applied project, Naval Postgraduate School). NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/17329
Cooper, R. G. (2017). Idea-to-launch gating systems: Better, faster, and more agile. *Research-Technology Management*, *60*(1), 48–52. https://doi.org/10.1080/08956308.2017.1255057
Dougherty, G. M. (2018). Promoting disruptive military innovation: Best practices for DoD experimentation and prototyping programs. *Defense Acquisition Research Journal*, *25*(1), 2–29.
Dubos, G., Saleh, J., & Braun, R. (2007). *Technology readiness level, schedule risk and slippage in spacecraft design: Data analysis and modeling*. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2007-6020
Eiband, M., Eveleigh, T. J., Holzer, T. H., & Sarkani, S. (2013). Reusing DoD legacy systems: Making the right choice. *Defense Acquisition Research Journal*, *20*, 154–173.
Etemadi, A. H., & Kamp, J. (2021a). Market and contractor factors affecting rapid acquisition strategies. *Systems Engineering*, *24*(4), 250–265. https://doi.org/10.1002/sys.21577
Etemadi, A. H., & Kamp, J. (2021b). Buying for the right battle: Determining defense acquisition strategies. *Proceedings of the 18th Annual Acquisition Research Program. Presented at the Acquisition Research Symposium*, 197–204. https://calhoun.nps.edu/handle/10945/68142
Flyvbjerg, B. (2006). From Nobel Prize to project management: Getting risks right. *Project Management Journal*, *37*, 5–15. https://doi.org/10.1177/875697280603700302
Fox, J. R. (2011). *Defense acquisition reform, 1960-2009: An elusive goal*. Center of Military History, U.S. Army. https://history.army.mil/html/books/051/51-3-1/index.html
Gansler, J. S., & Lucyshyn, W. (2010). National security acquisition challenges. *Strategic Studies Quarterly*, *4*(4), 13–31. https://www.airuniversity.af.edu/Portals/10/SSQ/documents/Volume-04_Issue-4/ganslerlucyshyn.pdf
GAO. (1992). *Weapons acquisition: A rare opportunity for lasting change* (NSIAD-93-15).
Garrett, R. K., Anderson, S., Baron, N. T., & Moreland, J.D. (2011). Managing the interstitials, a System of Systems framework suited for the Ballistic Missile Defense System. *Systems Engineering*, *14*, 87–109. https://doi.org/10.1002/sys.20173
General Services Administration. (2019). *Federal acquisition regulations* [WWW document]. http://farsite.hill.af.mil/reghtml/regs/far2afmcfars/fardfars/far/Far1toc.htm#TopOfPage
Grant, R. (2016, June 24). The second offset. *Air Force Magazine*. https://www.airforcemag.com/article/the-second-offset/
Grau, D., Back, W. E., Mejia-Aguilar, G. (2017). Organizational-behavior influence on cost and schedule predictability. *Journal of Management in Engineering*, *33*(5), 4017027. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000542
Hsiao, Y., Chen, C., Guo, R., & Hu, K. (2017). First-mover strategy, resource capacity alignment, and new product performance: A framework for mediation and moderation effects. *R&D Management*, *47*(1), 75–87. https://doi.org/10.1111/radm.12157
Jaifer, R., Beauregard, Y., & Bhuiyan, N. (2020). New framework for effort and time drivers in aerospace product development projects. *Engineering Management Journal*, *33*(2), 1–20. https://doi.org/10.1080/10429247.2020.1772950
Jimenez, C. A., White, E. D., Brown, G. E., Ritschel, J. D., Lucas, B. M., & Seibel, M. J. (2016). Using pre-Milestone B data to predict schedule duration for defense acquisition programs. *Journal of Cost Analysis and Parametrics*, *9*(2), 112–126. https://doi.org/10.1080/1941658X.2016.1201024
Jovel, J., & Jain, R. (2009). Impact of identified causal factors to “System of Systems” integration complexity from a defense industry perspective. *Global Journal of Flexible Systems Management*, *10*, 45–54. https://doi.org/10.1007/BF03396571
Katz, D. R., Sarkani, S., Mazzuchi, T., & Conrow, E. H. (2015). The relationship of technology and design maturity to DoD weapon system cost change and schedule change during Engineering and

Manufacturing Development. *Systems Engineering*, *18*(1), 1–15. https://doi.org/10.1111/sys.21281
Klein Woolthuis, R., Lankhuizen, M., & Gilsing, V. (2005). A system failure framework for innovation policy design. *Technovation*, *25*(6), 609–619. https://doi.org/10.1016/j.technovation.2003.11.002
Lingens, B., Winterhalter, S., Krieg, L., & Gassmann, O. (2016). Archetypes and basic strategies of technology decisions. *Research-Technology Management*, *59*(2), 36–46. https://doi.org/10.1080/08956308.2015.1137192
Lord, E. M. (2020). *Operation of the Adaptive Acquisition Framework* (DoDI 5000.02). Department of Defense. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500002p.pdf
Mankins, J. C. (2009). Technology readiness and risk assessments: A new approach. *Acta Astronaut*, *65*(9–10), 1208–1215. https://doi.org/10.1016/j.actaastro.2009.03.059
Markham, S. K., & Lee, H. (2013). Product Development and Management Association's 2012 comparative performance assessment study. Journal of Product Innovation and Management, *30*(3), 408–429. https://doi.org/10.1111/jpim.12025
McCool, J. (2012). *Using the Weibull distribution: Reliability, modeling, and inference*. John Wiley & Sons.
Mortlock, R. F. (2019). *Acquisition strategy formulation: Evolutionary/incremental development approach* (NPS-PM-19-179). Naval Postgraduate School. https://dair.nps.edu/bitstream/123456789/2768/1/NPS-PM-19-179.pdf
National Defense Authorization Act for Fiscal Year 2016, Pub. L. 114–92 (2015). https://www.congress.gov/114/plaws/publ92/PLAW-114publ92.pdf
Oakley, S. S. (2020). *Defense acquisitions annual assessment drive to deliver capabilities faster increases importance of program knowledge and consistent data for oversight* (GAO-20-439). Government Accountability Office.
Patil, K., & Bhaduri, S. (2020). "Zero-error" versus "good-enough": Towards a "frugality" narrative for defence procurement policy. *Mind and Society*, *19*, 43–59. https://doi.org/10.1007/s11299-020-00223-7
Peters, W., Doskey, S., & Moreland, J. (2017). Technology maturity assessments and confidence intervals. *Systems Engineering*, *20*(2), 188–204. https://doi.org/10.1002/sys.21389
Roberts, B., Mazzuchi, T., & Sarkani, S. (2016). Engineered resilience for complex systems as a predictor for cost overruns. *Systems Engineering*, *19*(2), 111–132. https://doi.org/10.1002/sys.21339
Sargent, J. F. (2018). *Global research and development expenditures: Fact sheet* (CRS Report No. R44285). Congressional Research Service.
Sargent, J. F., & Gallo, M. E. (2021). *The global research and development landscape and implications for the Department of Defense* (CRS Report No. R45403). Congressional Research Service.
Swartz, S. M. (2008). Managerial perceptions of project stability. *Project Management Journal*, *39*(4), 17–32. https://doi.org/10.1002/pmj.20078
Tao, L., Wu, D., Liu, S., & Lambert, J. H. (2017). Schedule risk analysis for new-product development: The GERT method extended by a characteristic function. *Reliability Engineering & System Safety*, *167*, 464–473. https://doi.org/10.1016/j.ress.2017.06.010
Tate, D. M. (2016). *Acquisition cycle time: Defining the problem (Revised)* (IDA Document No. NS D-5762). Institute for Defense Analyses.
van Oorschot, K. E., Eling, K., & Langerak, F. (2018). Measuring the knowns to manage the unknown: How to choose the gate timing strategy in NPD projects. *Journal of Product Innovation Management*, *35*(2), 164–183. https://doi.org/10.1111/jpim.12383
Weber, K. M., & Rohracher, H. (2012). Legitimizing research, technology and innovation policies for transformative change: Combining insights from innovation systems and multi-level perspective in a comprehensive "failures" framework. *Research Policy*, *41*(6), 1037–1047. https://doi.org/10.1016/j.respol.2011.10.015
Washington Headquarters Services. (n.d.). *Selected acquisition reports*. Department of Defense. Retrieved March 30, 2022, from https://www.esd.whs.mil/FOIA/Reading-Room/Reading-Room-List_2/Selected_Acquisition_Reports/
Williams, T. (2005). Assessing and moving on from the dominant project management discourse in the light of project overruns. *IEEE Transactions on Engineering Management*, *52*(4), 497–508. https://doi.org/10.1109/TEM.2005.856572
Wirthlin, J. R. (2009). *Identifying enterprise leverage points in defense acquisition program performance*. Massachusetts Institute of Technology.

Build Back Better: The Reemergence of American Manufacturing – Is Easier Said Than Done Learning from Building the Defense Industrial Base

Symantha "Sam" Loflin—has an MS in Program Management and a certification in Advanced Acquisition Studies from the Naval Postgraduate School, where she was a contributor to the 18th Annual Acquisition Research Symposium. She also holds a BS in finance from the University of Houston. Sam is an Earned Value Management (EVM) Analyst at Defense Contract Management Agency and has over 20 years of acquisition experience supporting DoD, NASA, and the military services. She recently served as an acquisition program manager on the Coronavirus Task Force that focused on building the industrial base for personal protective equipment. Sam's career began at NASA, supporting the Space Shuttle, ISS, and the Constellation Programs in Houston. [symanthaloflin@gmail.com]

Abstract

The reemergence of American manufacturing is easier said than done. On March 11, 2021, the three-part Build Back Better (BBB) agenda to **rescue, recover, and rebuild the country** became law. The agenda included the injection of billions of dollars in funding to small businesses that would have a domino effect by strengthening the American manufacturing supply chains, sparking innovation, and creating economic stability. On November 19, 2021, the House of Representatives (H.R.5376, 2021) voted 220-213 for the Build Back Better Act (BBBA), which remains stalled in the Senate.[1] On November 15, 2021, the ***Infrastructure Investment and Jobs Act*** (IIJA) was signed into law, and it benefits Small Businesses and Manufacturing.[2] These acts are transformational change measures for guiding and streamlining to achieve economic growth and sustainment of domestic sources in America. At the onset, the "*delivery of performance [will be] at the speed of relevance*" (Mattis, 2018, p. 10). This paper analyzes past and current whole-of-government measures to determine the state of the Defense Industrial Base (DIB) for the reemergence of American Manufacturing.

Research Issue

How will the U.S. rescue, recover, and rebuild the country using American manufacturing at the speed of relevance?

Research Results Statement

It takes a whole-of-government approach to strengthening the Defense Industrial Base (DIB) with the federal procurement of Made in America products and the growth of small businesses to achieve economic growth and national security. Three suggestions for consideration. 1. Communication of Small Business and Manufacturing successes. 2. Monitor implementation of government-streamlining measures. 3. Create a central repository for Small Business that includes mentors and protégé partnership monitoring.[3] The results are clear that the federal measures enacted to rescue, recover, and rebuild America under the Biden administration are successfully moving forward at the speed of relevance.

---

[1] Build Back Better Act, H.R. 5376, 117th Congress (2021-2022).

[2] Infrastructure Investment and Jobs Act, 117th Cong., Rec. 3684 (2021)

[3] Department of Defense. (2020, November). Volume 1, Chapter 10: Advana – Common Enterprise Data Repository For The Department Of Defense

Introduction

Defense Production Act of 1950 and Small Business Act of 1953

The Defense Production Act (DPA) of 1950 to establish a defense mobilization infrastructure in response to the Korean War. As amended over 50 times, it gives the President of the United States (U.S.) the authorities to influence the domestic industrial base (Peters, 2020). It preserves the industrial base by expanding the production of good and services for national security. In February 2018, Recommendation 21 of the Section 809 panel's findings revealed the importance of leveraging Small Businesses innovative capabilities that enhance warfighting effectiveness and readiness that preserve the industrial base.[4] The findings emphasized the necessity for DOD to refocus on the 1953 Small Business Act (SBA) that linked small business set-asides to the department's core mission of national defense. In addition, the findings included the lack of small business policies, industry outreach, and a logical business strategy. The panel concluded that the small business community provided innovative capabilities that are essential to national security by maintaining warfighting dominance and readiness.

Great Recession of 2008 and Inflation of 2022

President Biden's prior experience with the rescue, recovery, and rebuilding of America was instrumental in facilitating the end of the Great Recession of 2008. The theory and evidence as written by Roger E.A. Farmer, shows "…the stock market crash of 2008, triggered by a collapse in [home] prices that caused the Great Recession."[5] In January 2009, President Barack Obama's administration lead America into a recovery. On February 17, 2009, when President Obama signed the American Recovery and Reinvestment Act (ARRA) into law, he recognized Joe Biden's efforts in getting the legislation passed (White House, 2009a). The ARRA was a "major milestone on our road to recovery."[6] Soon after, Obama announced that Biden would oversee the implementation of the rescue, recovery, and rebuilding of America.[7]

There is a notable difference between rebuilding America from the Great Recession of 2008 and the current economy. America is in the "…deepest economic downturn…" since the Great Depression.[8] On December 21, 2021, the latest data inflation rate of 6.8% and the growth was at the highest rate in 39 years during a 12-month period [November 2020 – November 2021] (Weinstock, 2021b, pg. 1). Biden was successful with overseeing America's recovery through the ARRA and he is on track to continue that success with ***Build Back Better Act (BBBA)*** **and the** ***Infrastructure Investment and Jobs Act (IIJA)*** **that are** critical economic investments required to rescue America "…*at the speed of relevance*" (Mattis, 2018, p. 10). The following remarks by President Biden on delivering on "Made in America" commitments are as follows:[9]

---

[4] Report of the Advisory Panel on Streamlining and Codifying Acquisition Regulations, Volume 3 of 3. (2019, January). Section 809 Panel.

[5] The Stock Market Crash of 2008 caused the Great Recession: Theory and Evidence (Working Paper w17479).

[6] White House. (2009a, February 15). About the recovery.

[7] White House. (2009b, February 23). Vice President Biden to Oversee the Administration's Implementation of the Recovery Act's Provisions.

[8] Weinstock, Lida R. (2021a, May 11). Covid-19 and the U.S. Economy (CRS Report No. R46606). Congressional Research Service.

[9] White House (2022a, March 4). Remarks by President Biden On Delivering On Made In America Commitments.

"…when I say, **"Buy American,"** I mean buy all — **all American**. I want to increase the share of federal spending on goods and services that goes to **small businesses in America** — the backbone of our country…
Our **manufacturing** future, our **economic** future, our **solution**s to the climate crisis: They're all going to be **made in America**."

~ President Joseph R. Biden
REMARKS BY PRESIDENT BIDEN ON DELIVERING ON MADE IN AMERICA COMMITMENTS
March 4, 2022

Highlights of Building the Industrial Base

Buy American Act of 1933 and Buy American Act in 2021–2022

On March 3, 1933, during the Great Depression, Congress passed the Buy American Act (BAA), and President Hoover signed into law on his last day in office. When the BAA was enacted, it attempted "to protect domestic businesses and labor by establishing a price preference for domestic end products and construction materials in government acquisitions (Manuel, 2016, p. 1). In addition, the congressional oversight, by statute, requires agencies to submit a congressional report on procurement and compliance with the BAA that includes exceptions or trade agreement waivers. When solicitations contain the following clauses, federal government contracting officers who procure supplies are required to insert a FAR 52.225-2, Buy American Certificate and/or a FAR 52.225-6, Trade Agreements Certificate (TAA).

At the request of Senator Murphy, in December 2018, the Government Accountability Office (GAO) published their report on their review of four federal agencies implementation of the Buy American Act. The GAO reviewed 38 contracts from the Departments of Defense (DOD), Health and Human Services (HHS), Homeland Security (DHS), and Veterans Affairs (VA) and found that 6 contracts "…inaccurately recorded waiver or exception information" (Woods, 2018). The GAO found that steps should be taken by the Office of Management and Budget (OMB) to improve Buy American Act data and by the agencies to improve implementation guidance and training on the Act (Woods, 2018. pg. 1).[10]

Given the government's past BAA reporting compliance errors, the Biden-Harris administration is dedicated to improving the BAA through policies and laws that include the Federal Acquisition Regulation(s) (FAR), Executive Order(s) (E.O.), and the establishment of the first Made in America Office (MIAO). The following highlights are a few of the improvements related to the BAA:

On January 25, 2021, Executive Order 14005, Section 4 (a) and Section 7 are noteworthy measures, Section 4 (a): the Director of the Office of Management and Budget shall establish establish the Made in America Office within the OMB (White House, 2021a)[11] In April 2021, the Made in America Office (MIAO) opened to ensure "the future is made in America", strengthens domestic sourcing, and reduces the need for waivers. In addition, the office analyzes procurement waiver exceptions to "Made in America" laws and regulations that support United States manufacturing and domestic supply chains are allowable (GSA, 2021).[12]

[10] Woods, William T. (2018, December). Buy American Act: Actions Needed to Improve Exception and Waiver Reporting and Selected Agency Guidance (GAO-19-17), Government Accountability Office.

[11] White House (2021a, January 25). Executive Order 14005. Ensuring the Future Is Made in All of America by All of America's Workers was signed and Section 4 (a): Updating and Centralizing the Made in America Waiver Process.

[12] General Services Administration (2021). Made in America Office.

The MIAO reinforces government oversight and accountability of domestic procurement waiver exceptions shown as an extract from the official website in Figure 1.

Figure 1. Made in America Office - A Future Made in America (GSA, 2021)

Section 7: Supplier Scouting, establishes a significant role under the National Institute of Standards and Technology (NIST), Manufacturing Extension Partnership (MEP) National Network that shows partnerships committed to assisting **small** and **medium-sized** business in all 50 states and Puerto Rico.[13] In 2021, the MEP centers increased their assistance "…interacted with 34,307 manufacturers, leading to $14.4 billion in sales, $1.5 billion in cost savings, $5.2 billion in new client investments, and helped create or retain 125,746 jobs". A description of strengthening U.S. Manufacturing and the MEP is in Figure 2 (NIST, 2020).

E.O. 14005, Section 7: Supplier Scouting.

> "To the extent appropriate and consistent with applicable law, agencies shall partner with the Hollings Manufacturing Extension Partnership (MEP), discussed in the Manufacturing Extension Partnership Improvement Act (title V of Public Law 114-329), to conduct supplier scouting in order to identify American companies, including **small- and medium-sized companies**, that are able to produce goods, products, and materials in the United States that meet Federal procurement needs."(White House, 2021a)

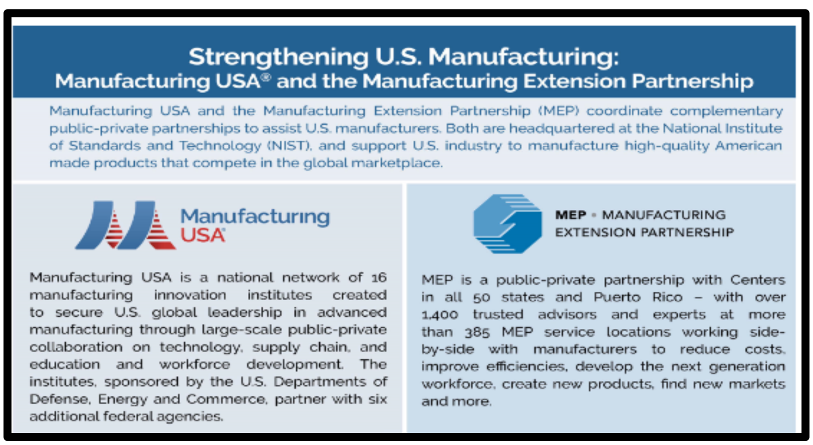


Figure 2. National Institute of Standards and Technology, Manufacturing Extension Partnership Infographic, 2020

---

[13] White House. (2021b, January 25). Executive Order 14005. Section 7: Supplier Scouting.

March 7, 2022, The U.S. Department of Defense (DOD), General Services Administration (GSA) and National Aeronautics and Space Administration (NASA), published the Federal Acquisition Regulation (FAR): Amendments to the Buy American Act Requirements, FAR Case 2021-008, was published with three notable changes listed below (FAR 25.101, 2022).[14] The rule and regulations for The Infrastructure Act requires that by Nov. 15, 2022, regulations will be implemented that amend the definitions of "domestic end product" and "domestic construction material". Following this announcement, International Association of Machinists and Aerospace Workers (IAMAW), President Robert Martinez Jr. released the following statement of gratitude on the IAMAW website:

This measure ensures that iron and steel products are made [to the greatest extent possible] with domestic components and provide a definition for *end product manufactured in the United States* as detailed below.

Figure 3. "Let's build the future right here in America" (Martinez, Jr., 2022)

1. **Increased Domestic Content Threshold**

   The current threshold is 55% for manufactured products purchased by the federal government. Table 1 describes the domestic content threshold timelines.

Table 1. Domestic Content Threshold Timelines (2022)

Effective Dates	Domestic Content Threshold
Now – Oct. 24, 2022	55%
Oct. 25, 2022 – Dec. 31, 2023	60%
Jan. 1, 2024 – Dec. 31, 2028	65%
Jan. 31, 2029, and after	75%

[14] FAR 25.101, General (2022, March 7). Amendments to the Buy American Act Requirement

2. **Exception for a Lower Domestic Content Threshold Due to Unavailability or Unreasonable Cost – "Fallback Threshold"**

   This change allows an agency to use the current 55% threshold for end products or construction materials when there is an absence of these materials that meet the new domestic content threshold, or the cost is deemed unreasonable.

3. **Increasing Price Preference for "Critical Items" and "Critical Components"**

   This mandates the application of a higher price preference for critical items and components in accordance with the 2021 E.O. 14017, "America's Supply Chains."

Small Business Improvement Acts

On February 3, 2022, Small Business Committee Passes and Recommends Five Bills to the House of Representatives that will help American small business entrepreneurs succeed, which are shown below.[15]

H.R. 6445: *"Small Business Development Centers Improvement Act of 2022" - to amend the Small Business Act to require an annual report on entrepreneurial development programs, and for other purposes.*

H.R. 6441: *"Women's Business Centers Improvement Act of 2022" - to amend the Small Business Act to improve the women's business center program, and for other purposes.*

H.R. 6450: *"SCORE for Small Business Act of 2022" - To amend the Small Business Act to reauthorize the SCORE program, and for other purposes.*

H.R. 4877: *"One Stop Shop for Small Business Compliance Act of 2021" - To amend the Small Business Act to require the Small Business and Agriculture Regulatory Enforcement Ombudsman to create a centralized website for compliance guides, and for other purposes.*

H.R. 6454: *"Small Business Advocacy Improvements Act of 2022" - To clarify the primary functions and duties of the Office of Advocacy of the Small Business Administration, and for other purposes.*

Office of Small Business Programs (OSBP) – DoD Mentor Protégé Program (MPP)

The OSBP is under the Small Business Act that established mandatory small business contracting goals and programs that apply to DOD and all Federal agencies (Office of Small Business Programs, 2022, April 5).[16] Its mission is to contribute to national security by maximizing opportunities for small businesses that provide combat supplies for our troops and economic sustainment for our nation. One of the highest responsibilities is the management of the DOD Mentor Protégé Program (MPP). The program is critical to developing high priority sectors of the DOD Industrial base (OSBP, 2022). A representation of some of the DOD MPP Project Spectrum Program Partnerships is shown in Figure 4 (Diaz, 2021, pg. 6).

---

[15] House Small Business Committee Republicans (2022, February 3). Small Business Committee Passes and Recommends Five Bills to the House of Representatives.

[16] Office of Small Business Programs (2022, April 5). Mentor-Protégé Program (MPP).

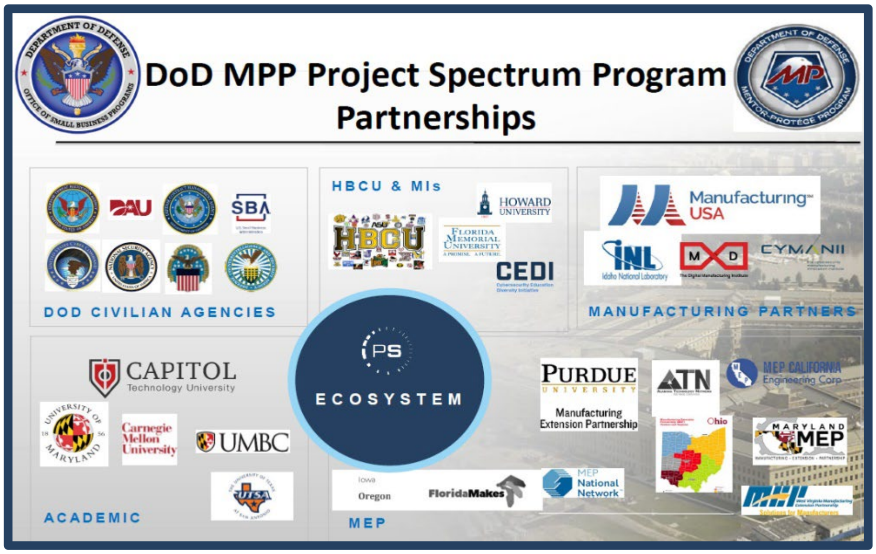


Figure 4. Extracted from DOD MPP Project Spectrum Program Partnerships (Diaz, 2021, p. 6).

On November 5, 1990, H.R. 4739 – National Defense Authorization Act (NDAA) for Fiscal Year (FY) 1991, directs the Secretary of Defense to establish a Mentor-Protege Program [MPP] in order to provide incentives to major DOD contractors (mentors) to help disadvantaged small businesses (protégés) perform as subcontractors and suppliers under DOD and other government contracts.[17]

On October 1, 1991, the DOD MPP was the first operative federal mentor-protégé program that since its inception as a pilot program. It has received continuous funding extensions as a pilot in spite of the 1994-scheduled expiration. Currently, it is funded through FY2026 for reimbursement of cost incurred under existing agreements and FY2024 for the formation of new agreements. DOD's MPP is the only federal pilot program that is mandated by law and receives authorized and appropriated funds (Mentor Protégé Pilot Program, 1990).[18]

Historically, the DoD's Mentor-Protégé Program is a front-runner with mentors' commitment to leveraging small business protégés in successfully growing the DIB, but the Office of the Secretary of Defense (OSD) must champion consistent support and funding for the MPP. In FY20, the MPP experienced a zeroed-out funding from the DoD in the FY2020 Defense Wide Review (DWR). The President's Budget Request (PBR) rescued funding for the MPP by adding it back in for FY2021 (Defense Business Board (DBB), 2022, p. 33). The notable facts in the DBB's (2022) MPP assessment for FY2021 are the **positive impacts of the MPP** and the challenges/recommendations in number 6: **Permanency of the MPP** shown in Figure 5 (2022, pgs. 71 and 83).[19]

---

[17] National Defense Authorization Act (1990, November 5). Pub. L. No. 101-510, 104 STAT. 1490, Title VIII: Acquisition Policy, Acquisition Management, and Related Matters - Part D: Miscellaneous, Sec. 831. Mentor-protégé pilot program (1990).

[18] Mentor-Protégé Pilot Program, Section 807 (a) of Pub. L. 102-484 (1991).

[19] Defense Business Board (2022, March 8). FY2022 Assessment of The Department Of Defense Mentor-Protégé Program.

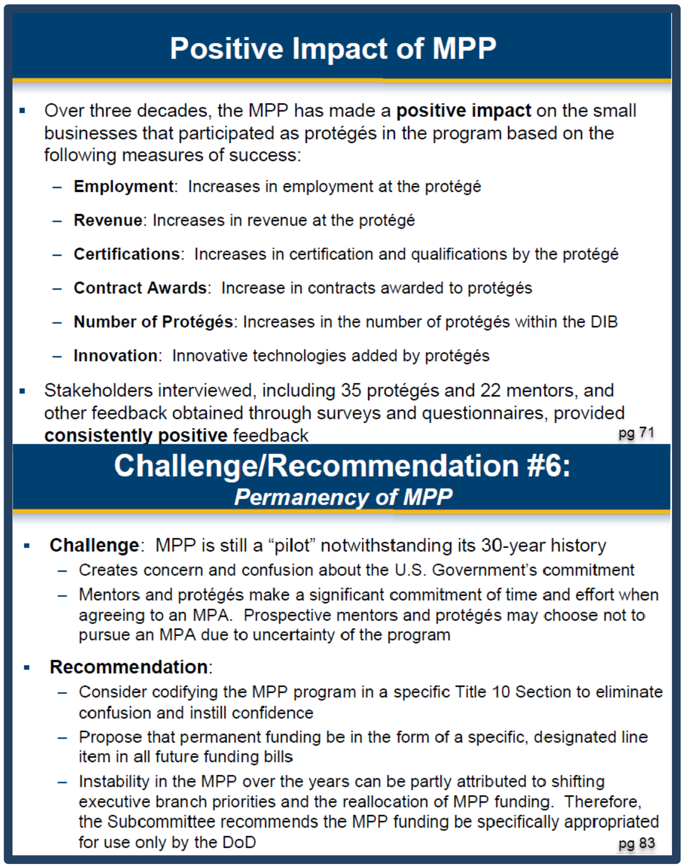


Figure 5. Adapted from the FY22-01 Assessment of The Department of Defense Mentor-Protégé Program (DBB, 2022).

References

American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, 123 STAT. 115 (2009). https://www.congress.gov/bill/111th-congress/house-bill/1/text

Build Back Better Act of 2021, H.R. 5376, 117th Cong. (2021–2022). https://www.congress.gov/bill/117th-congress/house-bill/5376/text

Congressional Research Service. (2016, April 26). *The Buy American Act—Preferences for “domestic” supplies: In brief* (R43140). https://sgp.fas.org/crs/misc/R43140.pdf

Congressional Research Service. (2020, March 2). *The Defense Production Act of 1950: History, authorities, and considerations for Congress.* https://crsreports.congress.gov/product/pdf/R/R43767

Defense Business Board. (2022, March 8). *FY2022 assessment of the Department of Defense Mentor–Protégé Program*. https://dbb.defense.gov/Portals/35/Documents/Meetings/2022/Feb%2022%202022/DBB%20FY22-01%20MPP%20Assessment%20Final%20Report%20w_Cover%20-%20Cleared.pdf?ver=s1EXfZ7pqYtqa5FAoDMwwg%3d%3d

Department of Defense. (2020, November). *Volume 1, Chapter 10: Advana—Common enterprise data repository for the Department of Defense* (DOD 7000.14-R). https://comptroller.defense.gov/Portals/45/documents/fmr/archive/01arch/01_10_Nov20.pdf

Department of Defense, Office of Small Business Programs. (2022, April 5). *Mentor–Protégé Program (MPP)*. https://business.defense.gov/Programs/Mentor-Protege-Program/

Diaz, K. (2021, May). *DoD Mentor-Protégé Program*. Department of Defense, Office of Small Business Programs. https://www.mda.mil/global/documents/pdf/Diaz,%20K%20-%20DoD%20Mentor%20Protege%20Program_MDA%20SB%20CONF%20(May_2021)FINAL.pdf

Executive Order 14017. (2021, January 25). *Ensuring the Future Is Made in All of America by All of America’s Workers.* https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/25/executive-order-on-ensuring-the-future-is-made-in-all-of-america-by-all-of-americas-workers/

FAR 25.101, General (2022, March 7). https://www.acquisition.gov/far/25.101

FAR 52.225-2, Buy American Certificate (2022, January 30). https://www.acquisition.gov/far/52.225-2

FAR 52.225-6, Trade Agreements Certificate (2021, February). https://farclause.com/FARregulation/Clause/FAR52.225-6_Basic-trade-agreements-certificate#gsc.tab=0

Farmer, R. E. A. (2011, October). *The Stock Market Crash of 2008 caused the Great Recession: Theory and evidence* (Working Paper w17479). National Bureau of Economic Research. https://safe.menlosecurity.com/doc/docview/viewer/docN19E029ECB435a5660aa793e930a409e61e7e4139e931205bedbb8954d1952804d013b26aace1
Federal Register. (2021a, January 19). *Maximizing use of American-made goods, products, and materials* (FAR 25.003). https://www.federalregister.gov/documents/2021/01/19/2021-00710/federal-acquisition-regulation-maximizing-use-of-american-made-goods-products-and-materials
Federal Register. (2021b, January 28). *All of America's workers make ensuring the future in all of America* (Executive Order 14005). https://www.federalregister.gov/documents/2021/01/28/2021-02038/ensuring-the-future-is-made-in-all-of-america-by-all-of-americas-workers
Federal Register. (2021c, February 24). *America's supply chains* (Executive Order 14017). https://www.federalregister.gov/documents/2021/03/01/2021-04280/americas-supply-chains
Federal Register. (2022, March 7). *General, subpart 25.1 – Buy American - Supplies* (FAR 25.101). https://www.federalregister.gov/documents/2022/03/07/2022-04173/federal-acquisition-regulation-amendments-to-the-far-buy-american-act-requirements
General Services Administration. (2021). *Made in America office*. https://www.madeinamerica.gov
House Small Business Committee Republicans. (2022, February 3). *Small Business Committee passes and recommends five bills to the House of Representatives*. https://republicans-smallbusiness.house.gov/news/documentsingle.aspx?DocumentID=403980
Infrastructure Investment and Jobs Act, Public Law 117-58. (2021, November 15). https://www.congress.gov/bill/117th-congress/house-bill/3684
Infrastructure Investment and Jobs Act of 2021, Pub. L. No. 117-58, 135 Stat. 429 (2021). https://www.congress.gov/bill/117th-congress/house-bill/3684/text
Manuel, K. M. (2016, April 26). *The Buy American Act—Preferences for "domestic" supplies: In brief* (CRS Report No. R43140). Congressional Research Service. https://sgp.fas.org/crs/misc/R43140.pdf
Martinez, R., Jr. (2022, March 7). *"Buy American" gets a huge boost—A true "Union Man" in the White House—Fighting for jobs in St. Louis.* International Association of Machinists and Aerospace Workers. https://www.goiam.org/news/departments/hq/politics-and-legislation/machinists-on-the-hill/buy-american-gets-a-huge-boost-a-true-union-man-in-the-white-house-fighting-for-jobs-in-st-louis/
Mattis, J. (2018). *Summary of the 2018 national defense strategy of the United States of America Department of Defense*. https://DOD.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf
Mentor-Protégé Pilot Program, Section 807 (a) of Pub. L. 102-484 (1991). https://uscode.house.gov/view.xhtml?req=504+agency+procedures&f=treesort&num=238
National Defense Authorization Act, Pub. L. No. 101-510, 104 Stat. 1490, Title VIII: Acquisition Policy, Acquisition Management, and Related Matters - Part D: Miscellaneous, Sec. 831. Mentor-protege pilot program (1990). https://www.congress.gov/101/statute/STATUTE-104/STATUTE-104-Pg1485.pdf
National Institute of Standards and Technology. (2020). *Manufacturing extension partnership infographics*. https://www.nist.gov/mep
Peters, H. (2020). *The Defense Production Act of 1950: History, authorities, and considerations for Congress* (CRS Report No. R43767). Congressional Research Service. https://crsreports.congress.gov/product/pdf/R/R43767
Section 809 Panel. (2018, January). *Report of the Advisory Panel on Streamlining and Codifying Acquisition Regulations, Volume 1 of 3*. https://discover.dtic.mil/wp-content/uploads/809-Panel-2019/Volume1/Recommendation_21.pdf
Weinstock, L. R. (2021a, May 11). *Covid-19 and the U.S. economy* (CRS Report No. R46606). Congressional Research Service. https://crsreports.congress.gov/product/pdf/R/R46606
Weinstock, L. R. (2021b, December 21). *Breaking down the U.S. inflation rate* (CRS Report No. IN11832). Congressional Research Service. https://crsreports.congress.gov/product/pdf/IN/IN11832
White House. (2009a, February 15). *About the recovery*. https://obamawhitehouse.archives.gov/blog/2009/02/14/a-major-milestone
White House. (2009b, February 23). *Vice President Biden to oversee the administration's implementation of the Recovery Act's provisions*. https://obamawhitehouse.archives.gov/realitycheck/the-press-office/vice-president-biden-oversee-administrations-implementation-recovery-acts-provision
White House. (2021a, January 25). *Ensuring the future is made in all of America by all of America's workers* (Executive Order 14005, Section 4(a)). https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/25/executive-order-on-ensuring-the-future-is-made-in-all-of-america-by-all-of-americas-workers/
White House. (2021b, January 25). *Supplier scouting* (Executive Order 14005, Section 7). https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/25/executive-order-on-ensuring-the-future-is-made-in-all-of-america-by-all-of-americas-workers/
White House. (2022a, March 4). *Remarks by President Biden on delivering on Made in America commitments.* https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/03/04/remarks-by-president-biden-on-delivering-on-made-in-america-commitments/
White House. (2022b, March 31). *Memorandum on Presidential Determination Pursuant to Section 303 of the Defense Production Act of 1950, as amended* (No. 2022-11). https://www.whitehouse.gov/briefing-room/presidential-actions/2022/03/31/memorandum-on-presidential-determination-pursuant-to-section-303-of-the-defense-production-act-of-1950-as-amended/
Woods, W. T. (2018, December). *Buy American Act: Actions needed to improve exception and waiver reporting and selected agency guidance* (GAO-19-17). GAO. https://www.gao.gov/assets/gao-19-17.pdf

The Price of Slavery: An Analysis of Human Trafficking Policy and Spending in Department of Defense Procurement

Capt. Willis C. Crouch IV, USAF—is a Contracting Officer (64P) in the U.S. Air Force. He commissioned in 2017 from Officer Training School after 14 years of enlisted service in security forces and contracting career fields.

1st Lt. Austin L. Morris, USAF—is a Contracting Officer (64P) in the U.S. Air Force. He started his career in 2018 at Buckley AFB, CO, after commissioning through AFROTC DET 017—Troy University, AL. He has a BS in kinesiology from Troy University, an MS in social science from Troy University, and an MBA from the Naval Postgraduate School.

1st Lt. Kevin P. Peaslee, USAF—is a Contracting Officer (64P) in the U.S. Air Force. KP has 18 years of military experience ranging from Security Forces, Inspector General, and Acquisitions. Additionally, KP has a master's in strategic purchasing, a master's in homeland security, a bachelor's in business administration, and an associate's in criminal justice.

Abstract

The Department of Defense (DoD) is charged with upholding the U.S. zero tolerance human trafficking policy in a world of complex, opaque supply chains and constrained human and capital resources. This study explores how the DoD can better leverage its acquisition workforce, sourcing expertise, and data to rigorously uphold the U.S. policy of zero tolerance for human trafficking in such an environment. As part of this analysis a program evaluation of the DoD's Combating Trafficking in Persons training was conducted for the acquisition workforce. This evaluation categorized the training related to prevention, monitoring, and response to human trafficking. A quantitative analysis of DoD spending was conducted to illuminate the amount of tax dollars spent in categories, based on product and service codes, where trafficking is most prevalent to help the DoD focus its efforts for program improvement. The spend analysis revealed areas that are most at risk. Globally, an estimated 24.9 million people are subjected to human trafficking, which generates an estimated $150 billion annually in illicit profits (White House, 2020, p. 8). Specifically, the DoD spent $13.1 billion (2018–2020) in countries that are not making significant efforts to combat human trafficking according to the Department of State. The research provides insight and recommendations on where the DoD should focus attention to address human trafficking in contracts constrained of prevention, monitoring, and response resources. Finally, we provided recommended courses of action to increase participation and enhance the mandatory training for the acquisition workforce.

Introduction

Human trafficking is an atrocity that all three researchers wanted to do something about. As contracting officers, we wanted to explore whether or not the United States could be spending tax dollars in areas that directly or indirectly support human trafficking.. We utilized spend analysis to explore gaps or areas of higher risk to human trafficking in government acquisitions given the U.S. government's (USG) zero tolerance policy. Our research question became:

> How can the DoD better leverage its acquisition workforce, sourcing expertise, and data to rigorously uphold U.S. policy of zero tolerance for human trafficking?

We were fortunate enough to be sponsored by the Department of Defense (DoD) Combatting Trafficking in Persons (CTIP) Program Management Office (PMO). The following paper will address the most critical areas of our research and the recommendations that were presented to the Department of State's (DoS) Senior Policy Operating Group (SPOG) in December 2021 and later to the Secretary of Defense Lloyd J. Austin III in January 2022. Our recommendations were also addressed by Deputy Secretary of Defense Kathleen Hicks to the President's Interagency Task Force (PITF) in January 2022.

Overview

"The Department of Defense (DoD) defines human trafficking, also known as trafficking in persons (TIP), as an abhorrent crime that is human rights abuse found in the forms of sex trafficking, forced labor, and child soldiering" (DoD Combating Trafficking in Persons Program Management Office, 2021, p. 1). Moreover, human trafficking falls within modern slavery; it is a moral, ethical, and legal problem that erodes the American foundation of "life, liberty, and the pursuit of happiness" (U.S. Declaration of Independence, 1776, para. 2).

The USG is charged as stewards of the taxpayers' dollar and the public's interest. Meaning, the USG cannot just throw aside the bedrock of moral, ethical, and legal decency of individual freedoms. As one of the largest spending organizations in the United States, the DoD is tasked to combat human trafficking through policy, action, and fiscal responsibility to further protect these freedoms and rights. The DoD has various measures in place to combat human trafficking. We specifically delved into FAR clause 52.222-50, Combating Trafficking in Persons. We discovered through this exploration that the FAR clause allows ambiguity by having an arbitrary $550,000 threshold and is limited to non-COTS service and commodities that are inside of the United States. These seemingly arbitrary policy triggers challenge the notion of a USG zero tolerance policy, and the USG does not appear to be maximizing its potential to prevent, monitor, and respond to human trafficking using contemporary data analytic methods. The study explores gaps in prevention, monitoring, and response (PMR) methods utilized by the DoD. We offer a first-of-its-kind use case of contemporary spend analytics and a nascent risk evaluation framework to better target areas of potential risk within DoD markets in an effort to align these methods with a true zero tolerance policy.

Prevention, Monitoring, and Response

We defined *prevention* in this research as methods employed to stop human trafficking from happening within federal acquisitions. We defined *monitoring* in this research as the ability to identify human trafficking within DoD acquisitions, through a standardized process. We defined *response* in this research as actions undertaken by those appropriately responsible when human trafficking is discovered or suspected within a DoD acquisition.

Preventative measures would be implemented in the pre-award/award stage, while monitoring and response measures would be implemented in the post-award/contractor performance stage. The research and recommendations aligned with strong prevention actions. Furthermore, strong prevention mechanisms could lead to less TIP and, therefore, lessen the burden of response.

Methodology

We used a mixed-method approach to conduct this research. We started the research with a quantitative spend analysis. We then conducted a qualitative program evaluation of the DoD's CTIP training for acquisitions and contracting. In 2019, the Office of Management and Budget (OMB) introduced a memorandum titled *Anti-Trafficking Risk Management Best Practices & Mitigation Considerations,* which urged federal agencies to conduct spend analyses in high-risk areas of human trafficking (Weichert, 2019). We used this memorandum as the motivation and conducted a spend analysis. We conducted scholarly literature and existing method research but were unable to find a USG or DoD TIP spend analysis framework.

We developed a spend analysis method given the limited researched area and lack of sufficient frameworks. Our spend analysis method quantified DoD spend, contracts, and contract actions in countries and specific goods or services that are tied to high-risk areas for human trafficking defined by the DoS Trafficking in Persons report. We used the Air Force Business Intelligence Tool (AFBIT) Lite application which is a product of the Air Force Installation Contracting Center Strategic Plans and Strategic Communications Division (AFICC/KA). AFBIT Lite is a visualization tool that gathers data from https://www.USASpending.gov and https://SAM.gov and displays it into targeted metrics used by the USAF contracting and category management communities. We identified our primary data elements and product service codes (PSC) using the General Services Administration (GSA) Fiscal Year (FY) 2020 PSC manual and the FSC PSC machine learning tool.

The qualitative portion of the research was a program evaluation of J3TA–US1328–C: CTIP DoD Acquisition Course. The program evaluation categorized the DoD Contracting and Acquisition CTIP computer-based training into the PMR elements that exist within contract management. The program evaluation further analyzed the acquisition CTIP training structure.

Results

We conducted a spend analysis for acquisitions during FY2019 and FY2020 in the markets for PPE, construction services, and food and food products based on the level of human trafficking tiers established by the DoS. We chose PPE, construction services, and food and food products because they were highlighted by various international agencies as having some of the highest levels of observed human trafficking. Table 1 illustrates the key insights from the spend analysis that we conducted.

Table 1. Synopsis of Spend Analysis Results

Results		
PPE	Construction	Food & Food Products
The DoD spent $6.66 billion for PPE in FY19 and FY20. **PPE is considered a COTS item**, meaning it is *not* covered under the criteria for having a CTIP compliance plan. **The DoD spent $5.26 million** on PPE **in Tier 3** countries.	The DoD spent $42.42 billion for construction services in FY19 and FY20. The DoD **spent $5.37 billion OCONUS on construction services**, which fall under the criteria for a defense contractor to submit a CTIP compliance plan. This reveals a **$37.3 billion gap** in CTIP prevention and monitoring since those construction services were performed domestically and not covered under the criteria for submitting a CTIP compliance plan.	The DoD spent $3.29 billion on food and food products in FY19 and FY20. In FY19 and FY20 there was **less than $1 million** spent for food and food products in Tier 3, Tier 2 Watch List, and Special Case countries. **Food and food products are considered COTS items** and do not require a CTIP contractor compliance plan.

Of the DoD's $66.52 billion in foreign spending during FY2018 through FY2020, $13.1 billion were in Special Case, Tier 2 Watch List, and Tier 3 countries. This indicates that nearly 20% of the DoD's foreign spending is at risk to human trafficking, given that it was spent in countries the DoS identifies as a higher risk. Furthermore, the analysis revealed that more than $161 million of the DoD's foreign spend was in Special Case countries from FY2018 to FY2020.

Personal Protective Equipment (PPE)

The DoD spent a total of $6.66 billion for PPE in FY2019 and FY2020. PPE is considered a COTS item, meaning it is not covered under the criteria for requiring a CTIP compliance plan from a defense contractor. Furthermore, a significant gap in CTIP prevention is created with $5.26 million spent on PPE in Tier 3 countries where there is no expectation for the countries to enforce anti-trafficking policies locally. We also note that, due to the dramatic increase for PPE created by the COVID-19 pandemic, a total of $6.66 billion was spent on PPE in just two fiscal years.

Construction Services

The DoD spent $42.42 billion for construction services in FY2019 and FY2020. In total the DoD spent $5.37 billion OCONUS on construction services. These are the only areas of construction spend that fall under the current regulatory criteria for a defense contractor to submit a CTIP compliance plan. This reveals a $37.3 billion gap in CTIP prevention and monitoring for those construction services performed domestically and not covered under the criteria for submitting a CTIP compliance plan. Like PPE, this leaves a significant amount of DoD spend at risk to human trafficking given that the construction services market is listed as an area of high risk by the Department of Labor.

Food and Food Products

For FY2019 and FY2020 there was less than $1 million spent for food and food products in Tier 3, Tier 2 Watch List, and Special Case countries. In total, the DoD spent $3.29 billion on food and food products in FY2019 and FY2020. In conclusion, like PPE, food and food products are considered COTS items and do not require a CTIP contractor compliance plan according to FAR 22.17 and FAR 52.222–50. Food products are highly heterogenous and involve diverse and complex supply chains. This area of spend would require further, more granular exploration to identify the highest areas of risk within the category.

Program Evaluation

In addition to a spend analysis, we conducted a program evaluation of the DoD's CTIP training for acquisitions personnel. The program evaluation categorized elements of the training related to PMR and human trafficking. The DoD CTIP training had a total of eight pages. Only seven of the pages contained content that could be evaluated. The pages were large in content, and some contained information that fit into multiple categories. The purpose of the program evaluation was not to critique the content of the training but to explore and illuminate how much, if any, of PMR attributes were covered within the training. The final page was a conclusion slide and was not applicable to categorization. The reviewers of the training are the authors of this research, and the results appear in Table 2. We conclude that the DoD CTIP training focuses heavily on the preventative aspect of the problem and is far more sparse on monitoring and response sections.

Table 2. Analysis of DoD CTIP Training

	Categories								
	Reviewer #1			Reviewer #2			Reviewer #3		
Pages	Prevention	Monitoring	Response	Prevention	Monitoring	Response	Prevention	Monitoring	Response
Page 1 of 8	X			X			X		
Page 2 of 8	X			X	X		X		
Page 3 of 8	X			X			X		
Page 4 of 8	X	X		X	X		X		
Page 5 of 8		X			X			X	
Page 6 of 8			X	X		X			X
Page 7 of 8	X	X		X			X		
Page 8 of 8	N/A			N/A			N/A		

Recommendations

The following recommendations were developed in response to our findings. First and foremost, these recommendations should be implemented in areas deemed to be high risk that are identified by a spend analysis. The recommendations have been grouped according to the PMR elements described earlier and should be implemented in their respective contract life cycle stages. The following paragraphs summarize a few of our prevention recommendations. In-depth analysis and the full list of recommendations can be found in our thesis at the Acquisition Research Program Defense Acquisition Innovation Repository (https://nps.edu/web/acqnresearch/dair).

Prevention

A recent DoD Inspector General (DoDIG) inspection revealed that there were contracts missing the mandatory clause, 52.222-50, Combatting Trafficking in Persons (DoDIG, 2019, p. i). FAR 22.1705(a)(1) mandates the CTIP clause to be in all solicitations and contracts. Thus, we recommended automating the inclusion of the clause in all DoD contract writing systems. An automation of the CTIP clause would essentially eliminate any solicitations and contracts missing the required clause. The automation prevents a contracting agency from not putting the CTIP clause in a solicitation or contract through a DoD contract writing system further strengthening zero tolerance to human trafficking.

DoD Instruction 2200.01 requires that acquisition-specific human trafficking be taken every three years by all acquisition personnel. We recommend before use and at time of log on to a contract writing system that personnel must be current in acquisition CTIP training. For example, a contract specialist in the Air Force would not be able to log in to the Contracting Information Technology (CON-IT) system if their training was not current. We envisioned the process would be something like DoD cyber awareness training, where an Air Force network user who is not current on their training is not allowed full access rights to their computer until the training is complete. This raises CTIP training to a level commensurate with a zero tolerance policy. This recommendation should further incentivize acquisition personnel to get the training done, and force leadership engagement when mission essential acquisition activities are impacted due to personnel not having access to their systems.

We recommend that contracting agencies that deal with contracts in high-risk areas to human trafficking, whether by regional or categorical market, create a CTIP acquisition representative to monitor their unit's CTIP training, conduct spend analyses, and participate in multi-function teams as they relate to the contract life cycle. For example, the CTIP acquisition representative would participate in industry days, post-award conferences, and annual surveillances. This recommendation falls in line with the OMB's memorandum on Anti-trafficking Risk Management Best Practices & Mitigation Considerations (Weichert, 2019).

Monitoring

The monitoring recommendations related to the post-award and contractor performance stages of the acquisition process. Monitoring, as it relates to acquisitions, is not intended to go and look for human trafficking, but instead our intent is to refocus activities already in place to increase awareness and provide tools to identify high-risk areas.

The Contracting Officer Representative (COR) is the eyes and ears to the contracting officer. CORs have a large set of responsibilities that relate to requirements generation, contractor performance and payment, and ensuring that what the government receives is according to what is contracted. A COR completes these duties by interaction with contractor employees through labor checks, inspections, and progress checks. We recommend including questions related to human trafficking in their daily inspection checklists. The human trafficking related questions presented during these inspections and labor checks will create opportunities for potential victims to have a chance to speak with a COR and raise flags, and/or focus the attention of the COR on potential human trafficking situations within their contract portfolio.

We recommend a risk model to help contracting agencies identify where TIP risk exists based on agency/organizational spend and type of market (e.g., PPE, food and food services, construction services, etc.), as shown in Figure 1.

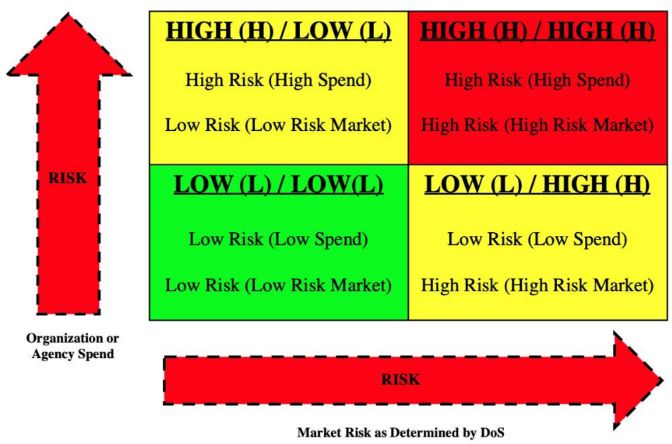


Figure 1. Risk Model Template

We developed the Human Trafficking Risk Dashboard Prototype (see Figure 2) in collaboration with the Air Force Installation Contracting Center, AFICC/KA. The Human Trafficking Risk Dashboard Prototype allows a user to conduct a spend analysis in seconds compared to the spend analysis we conducted for our research. The data is up to date with the most recent data received from FPDS-NG and cleansed by the AFICC/KA team. We use DoD spend data and combine it with the tier system from the DoS. This blend of data features allows agencies to instantly identify where their spend is going. For example, if an agency selected their DoDAAC and filtered to Tier 3 countries, they would be able to identify all their agency spend that is going to a high-risk area to human trafficking. Once identified, those contracts could be flagged for increased focus and awareness to human trafficking.

Combatant commands or high-level commands could run these exercises and work together with the DoD CTIP PMO to identify not only what they should be doing to ensure human trafficking is prevented, but also where to start. The tool is limited by areas where data is incomplete or unavailable. For example, DoD spend data is only captured at the prime level in this tool and only available at the first tier subcontract for some contracts. Much of the risk may lay at lower levels within the supply chain. However, this tool is a strong starting place from which to demonstrate proof of concept and encourage greater visibility into supply chain tiers an illuminating further risks.[130]

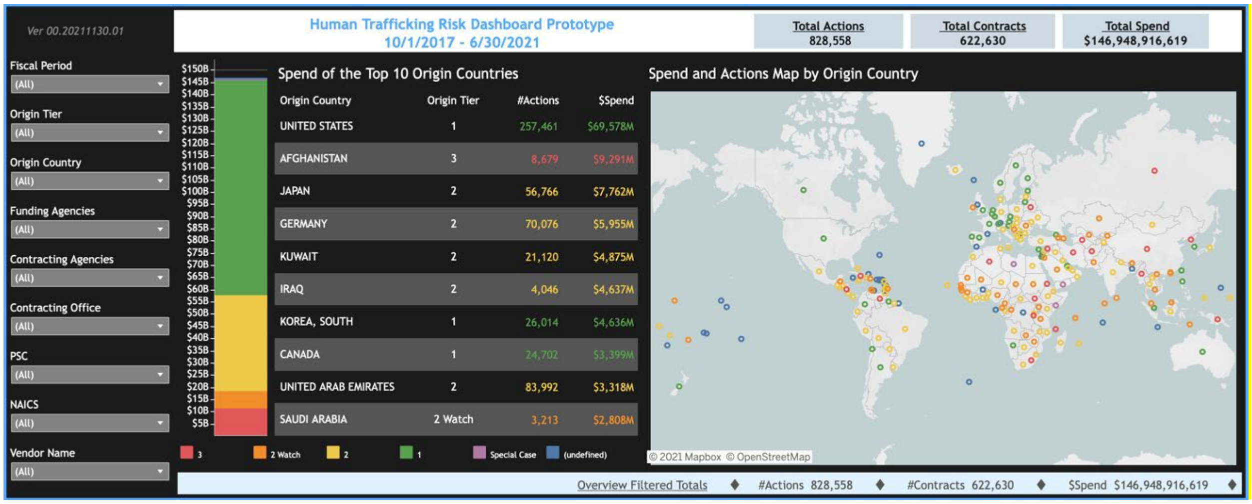


Figure 2. Human Trafficking Risk Dashboard Prototype Screenshot
AFICC/KA (2021, p.1)

Response

Unfortunately, human trafficking is a reality. We provided examples of TIP occurring in or around DoD contracts or acquisitions in our thesis. While acquisitions as a career field is not a primary stakeholder, like law enforcement, in response, they do play a role. We recommended acquisitions personnel have more interactions with agencies who are responsible for responding to human trafficking and when human trafficking is discovered around or within performance of a DoD contract. By bringing in acquisition employees could increase the buy-in, raise awareness, and strengthen response. To date our research has opened up channels of communication across the DoS; OMB; the Federal Law Enforcement Training Center; the Morale, Welfare, Recreation, and Resale Policy Directorate; and the President's Interagency Task Force to Monitor and Combat Trafficking in Persons (PITF). This is the best first step in improving our federal response and building a true zero tolerance environment.

Conclusion

Human trafficking is a huge global issue, and we would not be able to stop human trafficking with one thesis. We conducted research that would result in practical and tangible outcomes that could move toward the right direction and foster a true zero tolerance of human trafficking. The American taxpayer entrusts the USG to spend their dollars ethically. In order to

---

[130] These recommendations are not inclusive of all recommendations that were presented in our final research. Explanations fully capture the entirety of what was written on these recommendations in our research. However, if interested in both recommendations, please go to our thesis to find further information on them. The last recommendation section will discuss response recommendations.

do so we need to make sure that we are not allowing human trafficking to be affiliated with federal acquisition. Notwithstanding the fact that human trafficking is an atrocity for the victims, it also injects a supply chain risk by creating networks that are volatile, unreliable, and unethical. The USG cannot rely on supply chains that are associated with human trafficking and be fighting to rid human trafficking from the globe. With the USG's zero tolerance to human trafficking, the maximum effort needs to be put toward ensuring our acquisitions are solid in preventing, monitoring, and responding to human trafficking. If there are any further questions, please refer to our thesis or feel free to reach out to us. Thank you for your time and our united fight against human trafficking.

References

DoD Inspector General. (2019). Evaluation of DoD efforts to combat trafficking in persons in Kuwait. https://media.defense.gov/2019/Jun/20/2002147915/-1/-1/1/DoDIG-2019-088.pdf

DoD Combating Trafficking in Persons Program Management Office. (2021). *DoD combating trafficking in persons resources.* https://ctip.defense.gov/Resources/

U.S. Declaration of Independence. (1776). https://www.archives.gov/foundingdocs/declaration-transcript

Weichert, M. W. (2019, October 21). Anti-trafficking risk management best practices & mitigation considerations [Memorandum]. Office of Management and Budget. https://www.whitehouse.gov/wp-content/uploads/2019/10/M-20-01.pdf

White House. (2020). The national action plan to combat human trafficking. https://trumpwhitehouse.archives.gov/wp-content/uploads/2020/10/NAP-to-Combat-Human-Trafficking.pdf

An Investigation of the Role of System Effectiveness in the Acquisition and Sustainment of U.S. Defense Systems: 1958 to 2021

John M. Green—is a Senior Lecturer in the Systems Engineering Department, Naval Postgraduate School. His related experience includes employment as Manager of Advanced Systems at ITT Gilfllan and as a Principal Member of the Engineering Staff at Lockheed Martin. He holds an MBA and MS in computer science from University of New Haven, an MA in international relations from Salve Regina College, and a BS in physics from Saginaw Valley State University. He is also a graduate of the Naval War College, College of Command and Staff. He is currently a candidate for the PhD in systems engineering at Southern Methodist University.

Green is a Senior Member of IEEE and AIAA. He is a member of the Military Operations Research Society (MORS), the American Society of Naval Engineers, the Institute for Operations Research and Management Science, the Association of Old Crows, and the International Council on System Engineering. [jmgreen@nps.edu]

Abstract

This paper addresses the system effectiveness methodology and its intended role in acquiring and sustaining U.S. military weapon systems from 1958 to 2021. Given the prolonged period covered by this study and the many changes to the acquisition process, it would be reasonable to expect the methodology to change and adapt, and the study supports this assumption. The study uses the innovative approach of applying three qualitative methods: a structured review of the literature related to system effectiveness, a grounded theory analysis of the structured literature review, and a historiography of the initial grounded theory results. The research identifies five epochs, each marked by changes in the acquisition guidance. The conclusions are fourfold. First, the role of system effectiveness today is vastly diminished from its original purpose because the original material was not widely accessible to the community of interest during the formative years. The grounded theory result was that the concept was never allowed to mature because of changes that marked the second epoch's advent. Second, analysis of source documents provides insight into how to correct past misconceptions and incorporate system effectiveness into modern engineering. Third, the models developed in epoch one may have relevance for today's problems.

Introduction

This paper presents the outcome of an investigation into the role of system effectiveness in the acquisition and sustainment of U.S. defense systems from 1958 to 2021 (Green, 2022). The paper describes the inception of system effectiveness, the attempts to apply the concept, and ultimately, using an approach that combines a structured literature review, grounded theory analysis methods, and historiography techniques, a theory as to why system effectiveness fell into disfavor.

Given the prolonged period covered by this study and the many changes to the acquisition process that occurred during the period of interest, it would be reasonable to expect a change in the role of system effectiveness. The literature supports several changes throughout the time frame (Coppola, 1984). Still, changes were not necessarily driven by the acquisition process itself but by the underlying methodologies for developing systems that were also changing. The analysis shows a dynamic tension between the diverse communities involved in developing system effectiveness, which eventually led to the demise of the development of the concept as a methodology. However, it still exists as a concept in systems engineering texts, such as Habayeb's (1987) *Systems Effectiveness* and Wasson's (2015) *System Engineering Analysis, Design, and Development*.

Background

World War II highlighted the need for a concept by which the military could assess the effectiveness of weapons systems. Complex problems faced the Department of Defense (DoD). State-of-the-art solutions were required, as were methods by which to evaluate them. In the 1950s and 1960s, military systems were pushing the state of the art. Postwar systems were even more complex, encompassing programs such as the B-52 bomber and the Polaris missile program.

Moreover, given the nature of their missions, they had to be reliable and effective. As a result, Secretary of Defense McNamara introduced Systems Analysis into the defense acquisition process to address the quantification of cost and the effectiveness of weapons systems (Aziz, 1967). The initial response to McNamara's challenge came from the Army, Navy, Air Force, and analysis organizations within the DoD. As a result, throughout the 1960s, there was a flurry of activity by all three military services as they tried to incorporate McNamara's ideas into their vision of the acquisition process (Blanchard, 1967a).

A review of the literature related to System Effectiveness showed inconsistency in the concept from its first uses in the early 1960s through today. Earlier work by the reliability community started in the late 1950s. It served as the basis for developing the concept in the 1960s. While the literature shows little academic interest in the topic, a substantial body of work is available produced by the DoD and defense contractors. There was a serious effort by the DoD to develop System Effectiveness as a discipline highlighted by the development of the Weapon System Effectiveness Industry Advisory Committee (WSEIAC) methodology (Air Force Systems Command [AFSC], 1965b) to predict and measure System Effectiveness. By the early 1980s, the concept had all but disappeared from the literature. As System Effectiveness faded to a definition in the *Defense Acquisition Guidebook*, substantial interest in the topic of measures of effectiveness began to dominate the defense-related literature and is the most common concept currently in use. Further, the approach combines a structured literature review and grounded theory analysis. The structured literature review serves a unique role as the database for the grounded theory analysis.

Defining System Effectiveness: What Is System Effectiveness, and Why Is It Important?

Cost-effectiveness, System Effectiveness, integrated logistics support, and maintainability comprise the acquisition and sustainment process (Blanchard, 1967b). Of these four components, System Effectiveness is the linchpin. System Effectiveness is the starting point for deriving the other three components. As Blanchard (1967a) noted, "The ultimate goal of any system or equipment is to fulfill a particular mission for which it was designed. The degree of fulfillment is often referred to as System Effectiveness."

The original intent of system effectiveness was to focus management attention on overall effectiveness throughout the system life cycle. Further, system effectiveness is a framework for analytic methods to predict and measure the overall results of the analysis while placing the contributing characteristics in their proper perspective relative to the desired outcome of the system performing the mission. Thus, the system effectiveness framework provides a basis for developing needs and requirements during project definition and evaluating accomplishments during the acquisition and operation phases.

As developed by the DoD, System Effectiveness combined elements of reliability theory and system analysis. It was an outgrowth of work started in the 1950s by the reliability community and the system analysis work done by the RAND Corporation. Dordick noted in 1965 that it could be an uncomfortable relationship because the two groups viewed the problem from different perspectives.

The official definition of System Effectiveness is “A measure of the degree to which an item can be expected to achieve a set of specific mission requirements, and which may be expressed as a function of availability, dependability, and capability” (Blanchard, 1967a). However, the official definition does not indicate the scope and scale of system effectiveness as initially envisioned. Instead, system effectiveness represents an engineering management process concerned with describing, controlling, and measuring system performance in practice and a measure. Specifically, the management process provides a framework for system development through the four phases (conceptual, definition, acquisition, and operational) described by the *Air Force Systems Engineering Management Procedures Manual* (Gelbwaks, 1967).

As a measure, system effectiveness is one of the two elements of cost-effectiveness. Together, system effectiveness and cost-effectiveness represent the key elements of a 15-step management approach formulated to deal with the cost and complexity of modern military systems (AFSC, 1965a).

Restating the definition of System Effectiveness per Blanchard (1967a, 1967b), the management goal is to establish the probability that a system can successfully meet the operational demand within a given time when operated under specified conditions. This goal is the probability of success for the system. Accordingly, the framework focuses on evaluating or predicting the degree of effectiveness for any system configuration (existing or proposed). This degree of effectiveness has a cost associated with it that is the value used in the cost-effectiveness (CE) equation (Blanchard, 1967b),

$$CE = \frac{SE}{IC + SC}$$

where:

SE = cost of system effectiveness

IC = initial cost of procurement

SC = sustainment cost (life cycle cost)

System Effectiveness has three elements that determine both cost and the probability of success. This paper refers to them as the pillars upon which the System Effectiveness concept rests. These pillars are:

1. **Availability:** Is the system ready to perform its function?
2. **Dependability:** How well will the system perform during a mission?
3. **Capability:** Will the system produce the desired effects?

The first pillar is commonly referred to as operational availability or readiness, and the second pillar is commonly called mission reliability. Finally, some sources equate the third pillar, capability, with design adequacy, that is, is the design adequate for its intended mission? The three pillars are probabilities; thus, system effectiveness, the measure, is the product of availability, dependability, and capability. The intent was to use the System Effectiveness concept as a vehicle to proceed from predicted values in the conceptual phase of acquisition to empirical values as the system design matured and became operational and sustainment costs become paramount.

Statement of the Problem

Current literature referencing the system effectiveness concept (and, by extension, effectiveness measures) describes it ad hoc, based more on tribal lore than primary sources

(Reed & Fenwick, 2010). This approach is understandable because the legacy literature describes four system effectiveness models, one for the Army, one for the Air Force, and two for the Navy. In addition, terminology issues further exacerbate the problem. For example, the Navy has system effectiveness and operational effectiveness models. Further, the operational effectiveness model uses the three pillars (with different names), whereas the system effectiveness model has an entirely different approach that does not directly use the three pillars. The latter model is what the Navy intended to use for system effectiveness studies, even though the model was inconsistent with the Army or the Air Force (which use the three pillars but different names for the pillars). Finally, the Navy used the operational effectiveness model to train its analysts and supervisory personnel.

A second problem is the complexity of the mathematics used to describe system effectiveness. The common depiction of system effectiveness (the measure) is a scalar model of the three pillars' mathematical product. In reality, system effectiveness is the product of the availability vector [A] times the dependability vector [D] times the capability vector [C] (AFSC, 1965a), or

$$SE = [A][D][C]$$

A third problem is the lack of current references. The literature search turned up only one document written in the last 10 years that discussed system effectiveness: the *Operational Availability Handbook* (NAVSO P-7001; Assistant Secretary of the Navy, 2018). The handbook provided definitions and a computational approach to availability but left the determination of system effectiveness to the reader. The document illustrates ad hoc behavior by use of an incorrect definition of system effectiveness as follows: "Systems Effectiveness: The measure of the extent to which a system may be expected to achieve a set of specific mission requirements. It is a function of availability, **reliability**, dependability, **personnel**, and capability."

First, in the original model, system effectiveness is a function of availability, dependability, and capability for a reason. Moreover, the system effectiveness model answers the following questions: Is the system available when required? Is the system reliable throughout the mission? Furthermore, is the system capable of satisfactorily completing the mission? Second, the use of reliability and personnel is out of context with the intent of the original model. Reliability has a specific mathematical definition and is usually applied at the part level, whereas availability and dependability are system level measures. Finally, personnel would be an input parameter that impacts availability. As a result, the provided handbook definition does not support the system effectiveness criterion of being quantifiable and probabilistic (AFSC, 1965a, 1965b).

The final problem relates to the issue of measures of effectiveness and system effectiveness. AMCP 706-191 defined measures of effectiveness as an input into the system effectiveness process (Department of the Army, 1971). Measures of effectiveness became the ultimate measure with the demise of system effectiveness. Avoiding confusion between the two concepts is simple. First, system effectiveness is a function of the three pillars. Second, a measure of effectiveness measures how a system functions within its environment (Green, 2014). The difference between the two is a matter of context.

Specific Contribution of the Research

This paper reports the results of a study that thoroughly explores system effectiveness (Green, 2022). Specifically, this paper reports on the methodology used in that study. The unique contribution of this research is that it extends knowledge in the domain of system effectiveness related to acquisition and sustainment.[131] The research's value is its significant contribution to the system effectiveness body of knowledge. It presents a more current, thorough, and detailed analysis of a topic of interest to the acquisition and sustainment communities and supporting disciplines such as system engineering and reliability engineering. The research is novel because it uses several analysis techniques in a triangulated approach not generally applied to studies in this area. The research combines a structured literature review with grounded theory analysis and historiography techniques to develop a deeper and more detailed understanding of system effectiveness based on a comprehensive database of relevant papers from current and historical sources. This understanding provides a foundation for expanding the understanding and development of measures of effectiveness within the framework of system acquisition.

Related Work

Structured literature reviews and grounded theory have their roots in the social sciences. However, applications of grounded theory can extend beyond the social sciences. For example, Johnson recently published a doctoral dissertation titled *Complex Adaptive Systems of Systems: A Grounded Theory Approach* (Johnson, 2019; Johnson et al., 2018). In addition, structured literature reviews and grounded theory are being used in software engineering (Babar, 2019; Hoda, 2021; Stol et al., 2016).

Research Methodology

The research problem of investigating the role of system effectiveness in the acquisition process over 60 years does not fit into a traditional dissertation-like process. The answers to the research questions are qualitative, not quantitative. The data is the literature. Gathering and analyzing literature that went back before 1958 requires a different form of a literature review; hence, after some trial and error, the structured literature review concept was adopted for the subject research. Towards the end of the literature search, the need for a more detailed analysis process became apparent. The structured literature review was vital in determining the patterns in the literature. However, the structured literature review did not provide a methodology to aggregate the perceived patterns into a central concept or theory. Grounded theory methods were selected to meet this need because they focus on the topic at hand as limited by the researcher. Finally, assembly of the timeline–literature analysis concept led to the inclusion of historiography techniques to assist with developing the timeline.

There are four essential elements to developing a structured literature review and grounded theory analysis:

- **Step 1:** The research question
- **Step 2:** The structured literature review
- **Step 3:** The domain of inquiry
- **Step 4:** Critical elements in findings

The research into the combined or triangulated methods indicated five benefits (ResearchArticles.com, 2019):

---

[131] *Sustainment* is the appropriate term. *Sustainability* and *sustainable* have taken on specific meanings within the environmental community.

1) Increased validity of the results
2) A more nuanced view of the problem
3) Increased confidence in the results
4) Unique answers or results
5) A better understanding of the phenomenon involved

The techniques are sequential and recursive. Each pass through the data builds off the last pass, refining and distilling the observations into a central theme.

The Research Questions

The research question is the starting point, and the structured literature review represents the timeline and the data that support answering the research question. Finally, the domain of inquiry is the examination of the literature in the context of the timeline using grounded theory. The outcome is in the form of themes and patterns that emerge from the literature analysis with time.

The goal is an in-depth understanding of system effectiveness from its origins to 2021. The aim of the research is to assess how System Effectiveness evolved with changes in the U.S. DoD acquisition and sustainment processes. Table 1 presents the questions that served as drivers for the study.

Table 1. Research Questions

Question	Topic
Q1:	What factors led to the change in the role of System Effectiveness?
Q2:	What themes began to emerge with the changing role?
Q3:	What were noticeable patterns of change?

The Structured Literature Review

The structured literature review served two purposes in this study. First, the literature is the data, and using a search protocol identified material related to system effectiveness facilitating the development of an organized database. Second, the structured literature review served as the first filter in identifying potential patterns for the grounded theory analysis. Figure 1 describes the overall literature search process. The scoping study of Figure 1 identified possible sources to search. Table 2 presents the list of sources used. Also, the scoping study helped to limit the keywords used in the literature search. Table 3 lists prospective keywords developed from several sources, the primary source being the paper written by Tillman et al. (1978). Finally, Table 4 presents the final list used in the protocol.

Before undertaking the research, the Tillman et al. paper was a known entity. The paper surveyed the literature and identified 89 references specific to system effectiveness. The paper also described the main system effectiveness models developed to that point in time.

The focus of the search was on primary literature or original reports and secondary literature, which describes or summarizes the original writings. Also important is the category of the literature. What is its source? Table 5 presents the various literature categories used in the search. The order of search was (1) peer-reviewed material, (2) grey literature, and (3) books (texts and professional).

Grey literature is unpublished or not published commercially (see Table 5, items 2–7; Kamel, 2019). Because the development of system effectiveness was primarily a government effort, the majority of the literature retrieved fell into the grey category. The initial searches used different browsers and search engine combinations. For example, combining Firefox with DuckDuckGo or Google and Edge using Bing and Startpage was compared to Chrome using Google. The Google search engine was picked as the best option for this research because it had an excellent search string feature, and Google Scholar is a bonus. Additionally, the Chrome browser has a better download feature.

The literature retrieval process used three steps:

1) The use of a focused search string on the sources of Table 3
2) The use of “snowball” searches
3) A general web search using the focused search string

The use of a focused search string simplified the building of the database. Storing of the results was in folders named for the keywords. All filtering was manual, and sources identified but not available were not included in the database. Figure 2 uses “records” as a general term to cover papers, books, and reports.

TITLE-ABS ((“System Effectiveness” AND ((“keyword”))

The issue of using “system” vice “systems” is essential. It turns out that the use of systems provides lots of results, most of which are not usable. On the other hand, the use of “system” provides more focused results that are usable.

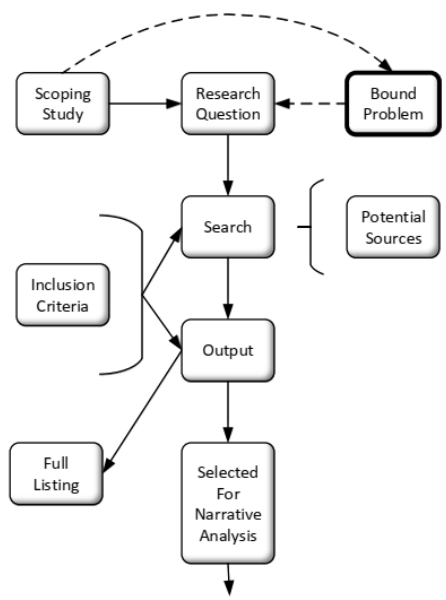


Figure 1. The High-Level Literature Search Process

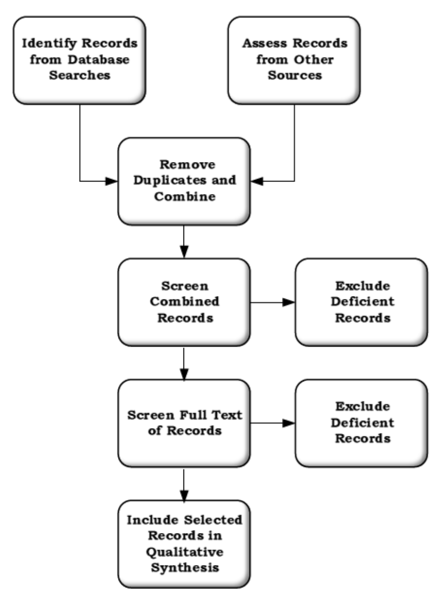


Figure 2. The Structured Literature Review

The desire to conduct as complete a search as possible drove the selection of sources to search (Table 2). Unfortunately, most 1950s and 1960s materials exist only in microfiche format, and COVID-19 restrictions limited access to archived materials. The search of Table 3 covered all sources listed. However, the primary focus was on the government column. The Defense Technical Information Center (DTIC) changed the public interface to use the Google search engine early in the research phase. This change had two undesired effects. First, the early searches were not repeatable, and the Google search engine provided few results. Fortunately, DTIC has a research portal that provides good results with the search string and the snowball search discussed below. Regretfully, the portal is not available to the public.

The second step was a “snowball” search (Wohlin, 2014) using the reference section of the selected papers. This search produced another 52 unique papers that were retrievable. Finally, the search string was also employed in a general web search, resulting in three conference proceedings found in Google Books unavailable from other sources. There were numerous references to conference proceedings as a significant source of information. However, few were available electronically, and those available were expensive.

Tables 3 and 4 present the list of keywords considered and selected, respectively. The keywords of Table 4 not used were tested but returned results not germane to system effectiveness. The primary focus of the search was thematic. What was the paper’s subject, and how did it relate to system effectiveness? The specific focus was on papers that addressed the theory, application, or programmatic issues.

The focus in examining search returns was title relevance, abstract relevance, and paper content, in that order. In addition, the search return had to demonstrate relevance to system effectiveness, the DoD, and the acquisition and sustainment process.

Table 2. Sources Used in the Literature Search

Academic	Government	Professional	Other
Scopus	Defense Technical Information Center - DTIC	Wiley	World Catalog
arXiv.org	Rand	IEEE Xplore	Proquest
ResearchGate	Naval Postgraduate School Calhoun Repository	Jstor	Internet Archive
Science Direct	National technical reports library - NTRL	SAE	SlideShare
Publons	Acquisition Research Journal	Operations Research	Google Scholar
Springer Science Plus Business	Web of Science	Library Genesis	Georgia Tech Research Library
Directory of Open Access Journals (DOAJ)	National Technical Information Service - NTIS	Naval Research Logistics Quarterly	
	MITRE	ARC(AIAA)	

Table 3. Prospective Key Words

Tillman, Hwang, and Kuo [11]	Other Sources
Reliability	Sustainment
Availability	Tactical availability
Operational readiness	Readiness
Repairability	Acquisition
Maintainability	Mission Reliability
Serviceability	Cost Effectiveness
Design adequacy	Operational Availability
Capability	Mission analysis
Dependability	Measures of effectiveness
Human performance	Measures of performance
Environmental effects	

Table 4. Selected Key Words

Reliability	Maintainability
Availability	Operational Availability
Operational readiness	Readiness
Dependability	Mission Reliability
Design adequacy	Cost Effectiveness
Capability	Mission analysis
Measures of effectiveness	Measures of performance

Grounded Theory and Coding the Data

McCall and Edwards (2021) identified three methodologies associated with grounded theory: the classic grounded theory of Glaser and Strauss (1967), the pragmatic grounded theory of Strauss and Corbin (1990, 1998), and constructivist grounded theory espoused by Charmez (2006, 2014). The discussion of the differences among these methodologies is beyond the scope of this paper. However, the study that this paper is reporting on used the pragmatic grounded theory approach. The following reasons are the basis for selecting this approach: First, it recognizes the literature as the phenomena to be studied. Second, it takes an interpretive approach that allows the development of a more profound understanding of the literature and the evolution of an abstract theory. Resultant theories are the researcher's interpretations of causal mechanisms. Third, the role of the researcher is that of an interpreter. However, this approach recognizes the researcher's personal experience and knowledge as a factor. The data sampling process is a back-and-forth effort that results in substantial memo writing and diagramming to identify and incorporate the data into manageable sets. The technique employs three distinct methods: open coding, axial coding, and selective coding. These sequential processes take the researcher through the steps to develop the data patterns (open coding) and examine the derived patterns for causality (axial coding). Axial coding confirms relationships between categories or bounds their applicability. Selective coding is about determining which category embodies the characteristics of the previously derived patterns. This category becomes the core category and represents the resulting theory. The overall procedure is recursive and proceeds until the sequence results in a candidate theory.

Table 5. Categories of Appropriate Literature

1. Peer-reviewed sources (journals and conferences)
2. Dissertations and theses
3. Professional journals
4. Conference proceedings (non-peer reviewed)
5. Government documents
6. Articles
7. Working papers and other unpublished material
8. Books

Analysis and Synthesis

The Data

Over 600 sources covering approximately 70 years (1950 to 2021) were the basis for developing the grounded theory. This research was unique in that the literature was the data. In addition, the resulting narrative was not linear. In the beginning, system effectiveness was the focus. However, in the end, the literature was more about analysis of alternatives (AoA), acquisition reform, and problems with reliability.

Step 1: The Analysis of the Data

Tables 6 and 7 are examples from the research report. Table 6 is the historiography, and Table 7 is the curated literature pertinent to the time frame. The aim was to present the main events during the period of interest with relevant documents published within the time frame. Comparing the event list with the publication list gives the reviewer an indication of what is of interest within the world of acquisition and sustainment during that period.

Table 6. Major Milestones, 1981–1990

Year	Milestone
1981	MIL-STD 721C published - Removed system effectiveness terminology
1981	Blanchard & Fabrycky, *Systems Engineering and Analysis* published
1983	DSMC publishes *System Engineering Management Guide* (SEMG) mentions system effectiveness
1984	TRADOC PAM 11-5 (COEA) revised with force focus
1985	MORS starts *Modular Command Evaluation Structure Study* (MCES)
1986	MCES report completed, First detailed study of Measures of Effectiveness
1986	DSMC (Arnold) publishes *Designing Defense Systems*
1987	Air Force R&M 2000 initiative
1987	Army reliability initiative coordinated with with Air Force R&M 2000
1987	Navy - Willoughby's "Best Practices Approach"

Table 7. System Effectiveness Publications, 1981–1990

Year	Title/Author
1980	System effectiveness models: an annotated bibliography (Tillman, Hwang, and Kuo)
1981	MIL-STD-721C, *Military Standard: Definitions of Terms for Reliability And Maintainability*
1981	MIL-HDBK 189 *Reliability Growth Management*
1982	DoDI 3235.1-H *Test and Evaluation of System Reliability, Availability, and Maintainability*
1983	The Measures of a System - Performance, Lifecycle-Cost, System Effectiveness, or What? (Blanchard)
1983	*System Engineering Management Guide*
1984	*The Human Operator and System Effectiveness* (Erickson)
1985	Design Adequacy: An Effectiveness Factor (Habayeb)
1985	Effectiveness Analysis of Evolving Systems (Karam)
1986	Command and Control Evaluation Workshop (Sweet, et al.)
1986	*Measures of Effectiveness in Systems Analysis and Human Factors* (Erickson)
1986	*Designing Defense Systems* (Arnold)
1987	Testing the Modular C2 Evaluation Structure and the Acquisition Process (Sweet/Lopez))
1987	*Systems Effectiveness* (Habayeb)
1990	*System Engineering Management Guide* (Kockler, et al.)

Table 8 presents the structured literature review's initial or open coding analysis.

Table 8. Initial Coding

Patterns
Changes with time
Changes with policy
Changes with DoD structure
Changes with technology
Changes with knowledge
Disparate technical disciplines
Tension among technical disciplines
Inconsistent models of System Effectiveness
Following fads
Lack of outside participation
Lack of participation by academia
Misuse of the concept
Lack of a consistent language

Step 2: Results of the Initial Coding

Initial coding is the search for trends and patterns in the database. The recursive analysis process initially divided the timeline into arbitrary 10-year increments. Further examination led to an initial division of the timeline into three epochs, defined as

1. McNamara's tenure as the secretary of defense
2. The introduction of the 5000 series of acquisition instructions in 1971
3. The advent of the Joint Capabilities Integration and Development System (JCIDS) process in 2002

Each pass through the data refined the timeline into sub-epochs that clarified the patterns and associated factors. The final result was five epochs. The adoption of commercial standards in 1993 and the current implementation of the Adaptive Acquisition Framework (AAF) complete the list. Table 9 lists the epochs with their causal event and interval of influence.

Table 9. EPOCHS

Year	EPOCH	Causal Event	Interval
1958	EPOCH I	Defense Reorganization Act	None
1971	EPOCH II	DoDD 5000.1	13 YRS
1993	EPOCH III	COTS	22 YRS
2002	EPOCH IV	JCIDS	9 YRS
2018	EPOCH V	Digital Engineering/AAF	16 YRS

The recursive process identified thirteen patterns. The grounded theory literature made it clear that behavior patterns were as crucial as a definable event. For example, a pattern of behavior might be the constant changing of personnel within a particular office in the DoD. A definable event might be the release of a new acquisition instruction. The two ideas merge when a new acquisition instruction is issued every time the leadership changes. This form of reasoning was the basic logic used for identifying the following patterns.

Initial Coding: Patterns and Concepts

1. **Changes with time:** The factors in this pattern address the history of system effectiveness as a function of time. It traces the development of the System Effectiveness models, their impact on military standards, and their subsequent input into the sustainment process. Example factors include the development of *reliability engineering*, *systems engineering* and *logistics engineering* alongside the attempts to develop system effectiveness.
2. **Changes with policy:** Identifies the significant policy changes that occurred to the acquisition structure with time from 1958 to the present. Sample factors include the *cycles of acquisition reform, the type of cycle,* and *the form of the changes*.
3. **Changes with DoD structure:** Factors in this pattern include *reorganization of research labs, changes in responsibility for system effectiveness within the DOD structure,* and *a lack of central authority*.
4. **Changes with technology:** The factors in this pattern refer to the emphasis of reliability over complexity. For example, the user community initially originally favored systems that demonstrated *mission reliability* over *capability.*
5. **Changes with knowledge and the knowledge base:** This could also read "changes with lack of knowledge or knowledge base." Factors include *loss of experienced analysts, inexperienced analysts, lack of reference material,* and *lack of example reports*. The latter two are problem areas because early material exists primarily as microfiche. Retrieval rates for the study was four of eight documents requested, and the waiting period was over 30 days.
6. **Disparate technical disciplines:** This pattern is distinguished by *a lack of common background or education.*

7. **Tension among technical disciplines:** Factors in this pattern include *a failure by some disciplines to see the big picture*. This is better known as "*if your only tool is a hammer, you tend to see problems as a nail."*
8. **Inconsistent models of system effectiveness:** The factors in this pattern center around *differences in what comprised effectiveness* and *similar terms that had different meanings among the various models of system effectiveness*.
9. **Following fads:** This pattern contains the factors that describe misguided attempts to redefine System Effectiveness to accommodate the management of the fad du jour. An example is *equating system effectiveness to quality at the expense of capability*.
10. **Lack of participation by industry:** This pattern is found throughout the literature. Factors include *proprietary methods that are time-tested* and *no financial incentives to change.*
11. **Lack of participation by academia:** Factors in this pattern address the *lack of research and publication* by the academic community.
12. **Misuse of the concept:** This is a common issue in the literature. Factors include *failure to understand the purpose of the system effectiveness concept* and *misrepresentation of the concept as solely a reliability model*.
13. **Lack of a consistent language:** Currently there is a lack of common and consistent terms for use when discussing system effectiveness. Factors include *no ontology and/or taxonomy* for system effectiveness and cost-effectiveness. The *lexicon* developed in the 1960s does not describe system effectiveness adequately. This has led to a confusion between what is system effectiveness and what is a measure of effectiveness.

Results of the Axial Coding

Axial coding is about causation. Again, this coding step is recursive, and the stopping point is when the grouping of patterns and their causal effects are complete. Table 10 presents the distillation of the 13 patterns into a list of candidates for the selective coding step. At this step, the outcome of the analysis was that five causal effects incorporated the 13 patterns.

Table 10. Axial Coding Results

Causal Effects
Tension among technical disciplines
Immaturity of the concept
Changes with time
Following fads
Changes with technology

Analyzing these five effects over the timeline leads to a sense of disarray. For example, Dordick (1965) identified the tension and lack of consistency between disciplines early on. The immaturity of the system effectiveness concept was a second contributor to the disarray. Too many people confused the concept with only reliability and maintainability (RAM) modeling because the various models shown in Blanchard's (1967a, 1967b) papers did not fully develop, nor were the models integrated into one consolidated model. In addition, McNamara was in office for only 4 years after he officially instituted system effectiveness. DoD Directive 5000.1 came 2 years later. Thus, taking the five causal effects together leads to the conclusion that the concept of system effectiveness was not allowed to mature. Development stopped, and people moved on.

The analysis of these five effects over the timeline leads to a sense of disarray and a lack of leadership. For example, the tension between disciplines was identified by Dordick (1965) at the beginning. The sense of disarray is heightened by the continued immaturity of the system effectiveness concept. Too many people confuse the concept with RAM modeling because the various models shown in the Blanchard papers (1967a, 1967b) were not fully developed nor integrated into one consolidated model. Three points stand out in the literature. First, Aziz (1967) pointed out the confusion in terminology and the lack of organized progress, particularly in performance analysis. Second, Coppola (1984) considered system effectiveness to be a transient idea and noted that system effectiveness gave way to life cycle cost as the emphasis. Third, the advent of MIL-STD-721C (DoD, 1981) supported Coppola's point, removing all references to system effectiveness.

The Theory

It would be easy to say, given the evidence, that system effectiveness is a failed concept, that the theory is one of failure. However, Habayeb (1987) presented a solid case to the contrary. The book presented three applications: hardware system evaluation, organizational development and evaluation, and conflict analysis. In addition, Rudwick (1969) identified three positive characteristics of the WSEIAC definition of system effectiveness:

1. The definition allows for the determination of the effectiveness of any system type.
2. The definition supports the measurement of any system in a hierarchy of systems.
3. The definition forces the analysis to focus on the three pillars.

Further, a search on Habayeb (1987) led to new material in Asia, specifically China. The Chinese have adopted the WSEIAC concept, referring to it as the ADC (for availability, dependability, and capability) model.[132] These points further support the theory that the development of system effectiveness stopped before maturity.

The Theory of Immaturity

The outcome of the selective coding step is the Theory of Immaturity. How can a concept that is in its sixties be immature? Simple. What may be signs of failure can also be signs that the idea never reached its full potential. That is the contention here. The literature shows that system effectiveness may have been a victim of a short attention span within the DoD environment. The era of system effectiveness began and ended with McNamara. Additionally, there were four variants of the system effectiveness model in play: one model for the Army, one for the Air Force, and two for the Navy (Blanchard, 1967a, 1967b). Four models for the same purpose do not indicate maturity. Finally, the services lost control of the acquisition process by the secretary of defense implementing DoD Directive 5000.1 in 1971. The literature indicates a lack of support by the disappearance of system effectiveness from DoD Directive 500.1 A mature process would most likely have received support.

Threats to the Validity of the Study

Research validity is essential in a study of this type where the result is subjective. Two factors drive the conversation: the literature review and the coding.

---

[132] The search used "Chinese and the WSEIAC model."

The challenge of the literature review is building a comprehensive database. In addition, there were negatives, such as the impact of COVID-19 isolation restricting access to physical materials. Nevertheless, despite the limitations, this research is a comprehensive study of system effectiveness with over 600 records.

Verification of the coding work occurred at each level of analysis. A researcher from another university experienced in grounded theory performed a confirmation analysis of the coding.

Finally, an evaluation criteria checklist presented as Table 11 guided the grounded theory analysis. The checklist also serves as a guide for the reader to follow the analysis results.

Table 11. Evaluation Criteria

1	How was the original sample selected? On what grounds?
2	What major categories emerged?
3	What were some of the events, incidents, actions, and so on that indicated some of these major categories?
4	Based on what categories did theoretical sampling proceed? That is, how do theoretical formulations guide some of the data collection? After carrying out the theoretical sample, how representative did these categories prove to be?
5	What were some of the hypotheses about relations among categories? On what grounds were they formulated and tested?
6	Were there instances when the hypothesis did not hold up against the observed? How are the discrepancies resolved? How did they affect the theory?
7	How and why was the core category selected? Was the selection sudden or gradual, difficult or easy? On what grounds were the final analytic decisions made?

Answering the Research Questions

Table 12 restates the research questions that this paper set out to answer. The initial coding identified thirteen factors that provide an answer to *Q1*. Chief among these factors is the tension between disciplines. The people involved practiced different disciplines and brought different perspectives and experiences to system effectiveness. Coppola was a reliability person, and his comment about system effectiveness meshes with Dordick's perspective about the difficulty in having different disciplines set aside their differences. The answer to *Q2* has three answers or themes. The first theme emphasized RAM at the expense of capability. The second theme was life-cycle cost (LCC), which incorporated the cost of RAM. Again, the capability pillar was not in the picture. The third theme focused on sustainment, which encapsulated the first two themes. It became more about a sustainable system than a capable system. The answer to *Q3* is yes. First, there was a shift in focus to LCC and, second, how to accomplish or perform analyses. The Cost and Operational Effectiveness Analysis (COEA) followed LCC and differed from a systems effectiveness study focused on the three pillars. The COEA followed a rigid, prescribed approach only to be replaced by the AoA concept, an analytical comparison of alternative material solutions that satisfy an established capability.

Table 12. Research Questions

Question	Topic
Q1:	What factors led to the change in the role of System Effectiveness?
Q2:	What themes began to emerge with the changing role?
Q3:	What were noticeable patterns of change involved?

The concept of system effectiveness is always lurking in the background, as exemplified by the *Operational Availability Handbook* (NAVSO P-7001) of May 2018. However, there are weaknesses in the concept. There is an issue with both a lexicon and a taxonomy. Thus, there is a need for an ontology to provide structure and organization. Resolution of these issues and needs would remove system effectiveness from "tribal lore" to established fact. The ontology would also provide a framework for the quantification of system effectiveness.

Summary of Research Results

Conclusions

The selected research method(s) served to clarify how system effectiveness came about, the attempts to make it viable, and how it meandered from the original concept. The triangulated approach led to the Theory of Immaturity by identifying patterns, concepts, and causal relationships. The research methods also clarified future research directions and highlighted issues and ideas that can improve the understanding and usage. The system effectiveness concept has application to a wide variety of systems engineering problems, including a system of systems architecture and cost-effective modeling with tools such as the Constructive Product Line Investment Model (COPLIMO).

Future Work

There are four recommendations: First, build the ontology. Second, refine the four system effectiveness models into one model. Third, establish the limits of the mathematical model. Finally, explicitly define the difference between systems. Finally, explicitly define the difference between system effectiveness and measures of effectiveness. Future work will develop an ontology and taxonomy that will provide a defined foundation to inform the application of system effectiveness and its methods. A second focus will be on developing case studies to illustrate the application of system effectiveness, clarify the lexicon, and uncover shortcomings not discussed in the literature.

Acknowledgement

The author would like to thank Professor Gary Langford, PhD, of Portland State University for his enthusiastic support, especially with the review of the coding of the data.

References

Air Force Systems Command. (1965a). *Weapon System Effectiveness Industry Advisory Committee (WSEIAC) final report of Task Group VI: Chairman's final report (integrated summary)*.

Air Force Systems Command. (1965b). *Weapon System Effectiveness Industry Advisory Committee (WSEIAC) final report of Task Group II: Prediction - measurement*.

Allen, G. H., DeMilia, R. M., & Gardella, C. E. (1967). Effectiveness consideration in applying AFSCM-375-5 requirements to control and surveillance systems. *SAE Transactions*, *75*(3), 490–504. https:/www.jstor.org/stable/44553422

Assistant Secretary of the Navy. (2018). *The operational availability handbook* (NAVSO P-7001). http://researcharticles.com/index.php/data-triangulation-in-qualitativeresearch/
Aziz. (1967, December). System effectiveness in the United States Navy. *Naval Engineers Journal*.
Babar, M. A., & Zhang, H. (2009). Systematic literature reviews in software engineering: Preliminary results from interviews with researchers. *3rd International Symposium on Empirical Software Engineering and Measurement*, 346–355.
Blanchard, B. S. (1967a, March). System worth, system effectiveness, integrated logistics support, and maintainability. *IEEE Transactions on Aerospace and Electronic Systems, AES-3*(2), 186–194.
Blanchard, B. S. (1967b, December). Cost effectiveness, system effectiveness, integrated logistics support, and maintainability. *IEEE Transactions on Reliability*, *R-163*(3).
Coppola, A. (1984, April). Reliability engineering of electronic equipment: A historical perspective. *IEEE Transactions on Reliability*, *R-33*(1), 29–35.
Department of the Army. (1971). *Engineering design handbook, system analysis and cost-effectiveness* (*AMCP-706-191*).
DoD. (1981, June 12). *Definition of terms for reliability and maintainability* (*MIL-STD-721C*).
Dordick, H. S. (1965, October). *An introduction to system effectiveness* (*Report No. P-3237*). RAND.
Gelbwaks, N. L. (1967). AFSCM as a methodology for system engineering. *IEEE Transactions on Systems, Science, and Cybernetics*, *SSC-3*(1), 6–10.
Green, J. (2014). *Establishing system measures of effectiveness*. AcqNotes. https://acqnotes.com/wp-content/uploads/2014/09/ Establishing-System Measures-of-Effectiveness -by-John-Green.pdf.
Green, J. (2022). *An investigation of the role of system effectiveness in the acquisition and sustainment of U.S. defense systems: 1958 to 2021* [Doctoral dissertation, unpublished manuscript].
Habayeb, A. R. (1987). *Systems effectiveness*. Pergamon.
Hoda, E. (2021, August). Socio-technical grounded theory for software engineering. *IEEE Transactions on Social Engineering*. https://arxiv.org/pdf/2103.14235.pdf
Johnson, B. (2019). *A framework for engineered complex adaptive systems of systems* [*Doctoral* dissertation, Naval Postgraduate School].
Johnson, B., Holness, K., Porter, W., & Hernandez, A. (2018). Complex adaptive systems of systems: A grounded theory approach. *The Grounded Theory Review*, *17*(1).
Kamel, F. K. (2019, July). The use of grey literature review as evidence for practitioners. *ACM SIGSOFT Software Engineering Notes*, *44*(3).
McCall, C., & Edwards, C. (2021). New perspectives for implementing grounded theory. *Studies in Engineering Education*, *1*(2), 93–107.
O'Brien, J., Remenyi, D., & Keaney, K. (2020, September 20). *Historiography: A neglected research method in business and management studies*. https://issuu.com/academic-conferences.org
Office of the Chief of Naval Operations. (2021, April 26). Operational availability of equipment and weapons systems (OPNAVINST 3000.12B). Department of the Navy.
Reed, R., & Fenwick, A. J. (2010). *A consistent multi-user framework for assessing system performance*. Cornell University. https://arxiv.org/abs/1011.2048
ResearchArticles.com. (2019, June 13). *Data triangulation in qualitative research*. http://researcharticles.com/index.php/data-triangulation-in-qualitative-research/
Rudwick, B. H. (1969). *Systems analysis for effective planning: Principles and cases*. Wiley.
Stol, K., Ralph, P., & Fitzgerald, B. (2016). Grounded theory in software research: A critical review and guidelines. *International Conference on Software Engineering Proceedings*.
Tillman, F. A., Hwang, C. L., & Kuo, W. (1978). System effectiveness models: An annotated bibliography. *IEEE Transactions on Reliability, R-29*, 295–304.
Wasson, C. S. (2015). *System engineering analysis, design, and development (2nd ed.*). Wiley.
Wohlin, C. (2014, May 13–14). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *EASE '14*, London, England, United Kingdom.

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net