SYM-AM-24-026

Proceedings
of the
Twenty-First Annual Acquisition Research Symposium

Thursday, May 9, 2024 Sessions
Volume II

Acquisition Research:
Creating Synergy for Informed Change

May 8–9, 2024

Published: May 1, 2024

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

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

To request defense acquisition research, please contact:

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
E: arp@nps.edu
www.acquisitionresearch.net

Copies of symposium proceedings and presentations; sponsored faculty and student research reports and posters may be printed from the **NPS Defense Acquisition & Innovation Repository** at **https://dair.nps.edu/**

Table of Contents

SYM-AM-24-026

Proceedings

of the

Twenty-First Annual Acquisition Research Symposium

Thursday, May 9, 2024 Sessions

Volume II

Acquisition Research:
Creating Synergy for Informed Change

May 8-9, 2024

Published: May 1, 2024

THIS PAGE INTENTIONALLY LEFT BLANK

WELCOME: RAYMOND D. JONES, COL, U.S. ARMY (RET), CHAIR, DEPARTMENT OF DEFENSE MANAGEMENT, NAVAL POSTGRADUATE SCHOOL

Raymond D. Jones, COL, USA (RET)—is the Chair, Department of Defense Management and Professor of the Practice at the Naval Postgraduate School. His last assignment in the Army was as the Deputy Program Executive Officer for the Joint Tactical Radio System (JTRS). Additionally, he served as the Military Deputy for the Director of Acquisition Resources and Analysis in the Office of the Under Secretary of Defense for Acquisition Technology and Logistics (USD(AT&L)), managed three Major Defense programs for the DoD in addition to his many operational and research and development assignments. He graduated from the U.S. Naval Test Pilot School in 1995 and is 1983 graduate of the United States Military Academy. He has a Bachelor of Science degree in Aerospace Engineering, a Master of Science Degree in Aeronautical Engineering from the Naval Postgraduate School, a Master's in Business Administration from Regis University, a Master's Degree in National Resource Strategy from the Industrial College of the Armed Forces and is currently a PhD candidate with the Graduate School of Information Sciences at the Naval Postgraduate School in Monterey California.

.

KEYNOTE SPEAKER: DR. WILLIAM A. LAPLANTE, UNDER SECRETARY OF DEFENSE FOR ACQUISITION AND SUSTAINMENT (A&S)

Honorable Dr. William A. LaPlante Senate, serves as the Under Secretary of Defense for Acquisition and Sustainment (USD(A&S)). In this role, he is responsible to the Secretary of Defense for all matters pertaining to acquisition; contract administration; logistics and materiel readiness; installations and environment; operational energy; nuclear, chemical, and biological defense; the acquisition workforce; and the defense industrial base.

Prior to this appointment, Dr. LaPlante served as President and Chief Executive Officer of Draper Laboratory, a research and development company specializing in advanced technology solutions in national security, space exploration, health care, and energy. Previously, he was senior vice president and general manager at MITRE National Security, where he oversaw the operation of two federally funded research and development centers and the U.S. Department of Commerce's National Institute of Standards and Technology.

Dr. LaPlante served as the Senate-confirmed Assistant Secretary of the Air Force for Acquisition, Technology, and Logistics from 2014 to 2017, where he aligned that Service's $43 billion acquisition enterprise budget with the Air Force vision and strategy. During his tenure, he forged a path forward on critical Air Force acquisition programs such as the B-21 long range strike bomber, while realizing nearly $6 billion in "should-cost" savings in other programs. Prior to this position, Dr. LaPlante spent 26 years at Johns Hopkins University Applied Physics Laboratory (APL), ultimately leading the Global Engagement Department where he was responsible for all of APL's work supporting offensive strike military capabilities. He also served as a member of the APL's Executive Council.

Dr. LaPlante has been a member of several scientific boards and commissions focused on maintaining national security, including the U.S. Strategic Command Senior Advisory Group, Naval Research Advisory Committee, and Defense Science Board. He joined other national experts as a commissioner on the congressionally-mandated Section 809 Panel, which performed a comprehensive review of Department of Defense acquisition policies and provided improvement recommendations, many of which became law.

Dr. LaPlante holds a doctorate in mechanical engineering from the Catholic University of America, a master's degree in applied physics from The Johns Hopkins University, and a bachelor's degree in engineering physics from the University of Illinois.

KEYNOTE SPEAKER: HONORABLE HEIDI SHYU, UNDER SECRETARY OF DEFENSE FOR RESEARCH AND ENGINEERING

Honorable Heidi Shyu—is the Under Secretary of Defense for Research and Engineering (OUSD(R&E)). In this role, she serves as the Chief Technology Officer for the Department of Defense (DoD), mandated with ensuring the technological superiority of the U.S. military, and is responsible for the research, development, and prototyping activities across the DoD enterprise. She also oversees the activities of the Defense Advanced Research Projects Agency (DARPA), the Missile Defense Agency (MDA), the DoD Laboratory and Engineering Center enterprise, and the Under Secretariat staff focused on developing advanced technology and capability for the U.S. military.

Previously, she served as the Assistant Secretary of the Army for Acquisition, Logistics and Technology (ASA (ALT)), from September 2012 to January 2016. Prior to this, she was Acting ASA (ALT) beginning in June 2011 and appointed the Principal Deputy in November 2010. As the ASA (ALT), she served as the Army Acquisition Executive, the Senior Procurement Executive, the Science Advisor to the Secretary of the Army, and the Army's Senior Research and Development official. She had principal responsibility for all Department of the Army matters related to logistics. Ms. Shyu also led the execution of the Army's acquisition function and the acquisition management system. Her responsibilities included providing oversight for the life cycle management and sustainment of Army weapons systems and equipment from research and development through test and evaluation, acquisition, logistics, fielding, and disposition.

Prior to her government service, Ms. Shyu was the Vice President of Technology Strategy for Raytheon Company's Space and Airborne Systems.

Ms. Shyu holds a Bachelor of Science in mathematics from the University of Brunswick in Canada, a Master of Science degree in mathematics from the University of Toronto, and a Master of Science degree in Electrical Engineering with a focus on System Sciences along with the Engineer's Degree from UCLA. She received an Honorary Doctorate of Science from the University of New Brunswick. She is also a graduate of the UCLA Executive Management Course Program.

A member of the Air Force Scientific Advisory Board from 2000 to 2010, she served as the Vice Chair from 2003 to 2005 and Chair from 2005 to 2008. Ms. Shyu is a member of the National Academy of Engineering and an American Institute of Aeronautics and Astronautics (AIAA) Honorary Fellow.

PANEL 14. SERVICE ACQUISITION FLAG OFFICER ROUNDTABLE

Thursday, May 9, 2024	
8:10 a.m. – 9:10 a.m.	**Chair: Michael Williamson, LTG USA (ret.)** President of Lockheed Martin International **Panelists:** **Major General Alice W. Trevino, USAF**, Deputy Assistant Secretary for Contracting, Office of the Assistant Secretary of the Air Force for Acquisition, Technology and Logistics **Brigadier General Frank J. Lozano, USA,** Program Executive Office, Missiles and Space **Brigadier General David C. Walsh, USMC,** Commander, Marine Corps Systems Command **Mr. Christopher P. Manning,** Deputy Assistant Secretary of the Army for Research & Technology (DASA R&T) and Army Chief Scientist

Michael Williamson, LTG USA (ret.)—is the president of Lockheed Martin International and senior vice president for Global Business Development & Strategy at Lockheed Martin Corporation. In this role, Williamson is focused on bringing integrated solutions to customers who rely on Lockheed Martin's capabilities and technologies to support their missions and address their most pressing needs. His responsibilities also include establishing comprehensive strategies across the enterprise that will enable future growth.

Previously, Williamson served as vice president and general manager for Lockheed Martin Missiles and Fire Control (MFC), where he was responsible for operational excellence, a diverse portfolio of products and business enabling initiatives.

He also previously served as vice president of Tactical and Strike Missiles for MFC. In this capacity, he managed significant programs in the areas of Hypersonic Weapon Systems, Close Combat Systems, Strike Systems, Precision Fires and Advanced Programs.

Williamson joined Lockheed Martin in 2017 following a distinguished career as a lieutenant general with the U.S. Army. He served as the principal military deputy to the assistant secretary of the Army for Acquisition, Logistics and Technology and director of Acquisition Career Management. He also served as a congressional fellow on Capitol Hill.

Williamson holds a bachelor’s degree in business administration from Husson University, a master’s in systems management from the Naval Postgraduate School, and a Ph.D. in business administration from Madison University. He is also a graduate of the Advanced Management Program at the Harvard Business School.

Major General Alice W. Treviño, USAF—serves as the Deputy Assistant Secretary for Contracting, Office of the Assistant Secretary of the Air Force for Acquisition, Technology and Logistics, the Pentagon, Arlington, Virginia. She is responsible for all aspects of contracting relating to the acquisition of weapons systems, logistics, operational and enterprise efforts for the Air Force and provides contingency contracting support to the geographic combatant commanders. She leads a highly skilled staff of mission-focused business leaders and acquisition change agents to deliver $825 billion in United States Air Force and Space Force platforms. Additionally, she is the Contracting Functional Manager for over 8,000 professionals, who execute programs worth $70 billion annually for the Department of the Air Force.

Maj. Gen. Treviño received her commission from the U.S. Air Force Academy in 1993 and is a joint qualified officer with extensive deployment experience in support of combat, humanitarian and peacekeeping/enforcement operations to Croatia, Turkey, Oman, Kuwait and Afghanistan.

Prior to her current assignment, she was the Commander of the Air Force Installation Contracting Center. Maj. Gen. Treviño has also served as the Deputy Secretary of Defense's Principal Military Assistant,; unlimited dollar warranted procuring Contracting Officer for major defense programs; and the Senior Contracting Official-Afghanistan for U.S. Central Command. She has commanded two Air Force units at the squadron level, joint units at the group and wing levels, and an Air Force unit at the wing level.

Brigadier General Frank J. Lozano, USA—is the Program Executive Officer (PEO), Missiles and Space, Redstone Arsenal, AL. He is responsible for the development, production, fielding, sustainment, and international program aspects for assigned missile and space systems. BG Lozano assumed his current position in August 2022.

BG Lozano was assessed into the Army Acquisition Corps in 2001 and graduated with an MBA from the University of Texas at Arlington. He served with Lockheed Martin Missiles and Fire Control in Grand Prairie, TX as part of the Training with Industry (TWI) program.

After completion of Command and General Staff College, BG Lozano was assigned as the Assistant Product Manager for Project Manager Soldier Weapons, PEO Soldier, followed by an assignment as an Ammunition and Demolition System Acquisition Manager for the Special Operations Command (SOCOM) and the Army Research Development and Engineering Command (RDECOM).

In 2008, BG Lozano was assigned as a Department of the Army System Coordinator (DASC) for Tactical Missile Systems and Ballistic Missile Defense Systems. BG Lozano was selected to be a Special Assistant for the Army's Vice Chief of Staff, GEN. As the Special Assistant, he provided insight, advice, and counsel on Army acquisition programs crossing many different functional capability areas.

BG Lozano commanded the Product Management Office for Soldier Protective Equipment, PEO Soldier from 2011 until 2014. Afterwards, he was assigned to the Joint Staff, J-8 Capabilities and Acquisition Division. Upon graduation from the US Army War College, BG Lozano was assigned as the Project Manager for the Lower Tier Project Office, PEO Missiles and Space from 2017 until 2020, followed by an assignment as the Integrated Fires and Rapid Capability Office PM. From April 2021 to May 2022 BG Lozano served as the ASA(ALT) Chief of Staff.

BG Lozano's operational and combat experience include deployments to Bosnia, Kuwait and Iraq. His awards and decorations include the Parachutist Badge, Ranger Tab, Legion of Merit, Bronze Star Medal, Joint Service Commendation Medal, the NATO Service Medal, the Army Staff Identification Badge, and the Joint Staff Identification Badge. He is certified in Program Management; Contracting; System Research; Planning and Engineering; and System Test career fields.

Brigadier General David C. Walsh, USMC— is the Commander, Marine Corps Systems Command, a native of Brooklyn, New York, was commissioned into the Marine Corps in August 1992 after graduating from the University of Virginia. He reported to Naval Air Station Pensacola for flight training and was designated a naval aviator in February 1995.

Brigadier General Walsh served with Marine Light Attack Helicopter Squadron-167 as an AH-1W pilot deploying with the 22nd Marine Expeditionary Unit. He attended the Marine Weapons and Tactics Instructor Course at Marine Aviation Weapons and Tactics Squadron-1, earned every AH-1W qualification possible, and served in various operations and maintenance billets.

Brigadier General Walsh later served as the H-1 lead test pilot for Air Test and Evaluation Squadron-31 and Weapons System Support Activity military deputy at Naval Air Weapons Station China Lake. He participated in developmental flight testing of the AH-1Z and UH-1Y, as well as multiple other flight test programs.

Brigadier General Walsh was selected in the first cadre of Marine Corps Acquisition Officers and has held numerous acquisition positions within the aviation community to include the H-1 assistant program manager for systems engineering and the director of operations at Fleet Readiness Center East, managing all depot-level operations including flight testing. He served as the Program Manager for Specialized and Proven Aircraft, responsible for the H-46 helicopter, Naval Test Pilot School aircraft, adversary aircraft, contracted air services, and multiple Foreign Military Sales cases. He then served as the Program Manager for the Marine Light/Attack Helicopter Program.

In March 2021, Brigadier General Walsh was selected as the Military Assistant to the Assistant Secretary of the Navy for Research, Development, and Acquisition. He was appointed as the acting Program Executive Officer Land Systems in February 2022 and assumed command of Marine Corps Systems Command in June 2022.

Brigadier General Walsh is a graduate of the Amphibious Warfare School and U.S. Naval Test Pilot School. He earned a master's degree in technical management from Johns Hopkins University and graduated from USAF Air War College. He is Level III acquisition certified in Program Management, Test and Evaluation and in Production, Quality and Manufacturing; and Level II certified in Engineering. Brigadier General Walsh has accumulated over 2,500 flight hours in more than 30 types of aircraft. His decorations include the Legion of Merit (three awards), Meritorious Service Medal (four awards), Navy Commendation Medal, and various campaign and unit awards.

Mr. Christopher P. Manning— serves as the Deputy Assistant Secretary of the Army for Research & Technology (DASA R&T) and Army Chief Scientist. He is responsible for policy and oversight of the Army's Research and Technology program. In this position, Mr. Manning is charged with identifying, developing, and demonstrating technology options that inform and enable effective and affordable capabilities for the Soldier. His science and technology (S&T) portfolio covers basic research to demonstrating component, subsystem, manufacturing technology, and technology system prototypes. The S&T portfolio is executed by the Army's research, development and engineering laboratories and centers; academia; and industrial and international partners. During his career, Mr. Manning has served in a variety of leadership and technical roles within the acquisition and research and development communities, including assignments with the U.S. Army Combat Capabilities Development Command (DEVCOM) Headquarters, Communications Electronics Research, Development and Engineering Center; Program Executive Office Command, Control, Communications – Tactical; Office of the Assistant Secretary of the Army (Acquisition, Logistics and Technology); and Program Executive Office Intelligence, Electronic Warfare, and Sensors. Most recently, Mr. Manning served as the Deputy Director for Force Development, Deputy Chief of Staff, G-8. With his experience in the research, development and engineering of experimental and state-of-the art capabilities for warfighter needs, Mr. Manning supports an ecosystem that further enriches Army research and development through partnerships with industry and academia, forged through the Army's Technology Transfer Program.

PANEL 15. INTERNATIONAL ACQUISITION: PARTNERS AND ADVERSARIES

Thursday, May 9, 2024	
10:30 a.m. – 11:45 a.m.	**Chair: Danial (Dino) Pick, Director,** International Graduate Programs, Naval Postgraduate School ***Bridging Sectors Over the Valley of Death: How DIANA's Dual-Use, Commercially Minded, and Process-Oriented Procurement Strategy Will Help Maintain NATO's Technological Edge*** Hope Hopkins, George Washington University ***Acquiring Technology for Allies and Partners*** Matthew Reed, Institute for Defense Analyses ***Resourcing Innovation that Registers: Sensing the PLA's Threat Perceptions around Major Defense Acquisition Programs*** Emily de La Bruyere, Horizon Advisory Nathan Picarsic, Horizon Advisory

Danial (Dino) Pick—Mr. Danial Pick serves as the Director of International Graduate Programs at the Naval Postgraduate School in Monterey, CA. Mr. Pick previously served as Principal Deputy Assistant Secretary of Defense for Special Operations and Low-Intensity Conflict (ASD (SOLIC)). In this capacity he assisted the ASD (SOLIC) in the execution of Department-wide counterterrorism and irregular warfare policy, operational oversight, and execution of service secretary-like responsibilities over U.S. SOCOM. Previously he served as the first Deputy Assistant Secretary of Defense for Special Operations Policy and Programs, a position called for in bipartisan legislation as a means of strengthening civilian oversight of U.S. Special Operations Forces. In this capacity he was responsible for advising the ASD (SOLIC) and the Secretary of Defense by formulating, recommending, integrating and implementing policies and strategies on matters pertaining to the organization, training and equipping of special operations forces including oversight of the $13.4B MFP-11 Special Operations budget. Mr. Pick graduated in 1987 from the University of Washington with a Bachelor of Arts degree in Near Eastern Languages and Civilization. He also holds a Master of Military Art and Studies from the Marine Corps University, and a Master of Arts in Near Eastern studies from Princeton University. He was a National Security Affairs Fellow at the Hoover Institution at Stanford University.Mr. Pick served for 29 years in the US Army as a military intelligence and Middle East foreign area officer. He served with the 1st Special Forces Group (Airborne) working with partners and allies throughout the Indo-Pacific region. His combat deployments included service with 3/66 Armor Battalion in Desert Storm in 1990-91, the 10th Special Forces Group (Airborne) during the invasion of Iraq in 2003, and service in Afghanistan in 2009. Mr. Pick helped develop the counterterrorism capabilities of the Iraqi and Jordanian Armed Forces while serving at the U.S. Embassy in Amman, Jordan, from 2003-2006, before serving as a Policy Officer in the Office of the Secretary of Defense from 2006-2008 building the partner capacity of the Iraqi Security Forces. Mr. Pick commanded the Defense Language Institute from 2010-2014 providing foreign language capability to special operations forces and the intelligence community. Mr. Pick has served as a municipal government leader in the Cities of Monterey and Del Rey Oaks, California. He currently serves as an elected official on the Monterey Peninsula Airport Board. He speaks Assyrian, Arabic, Farsi, and Dari. He is married to Karen Pick. They have two children, Dalton (24) and Lauren (21). Mr. Pick is a member of the Global SOF Foundation, the OSS Society, Chair of the Big Sur Marathon Foundation Board, and a member of the Rotary Club of Monterey.

Bridging Sectors Over the Valley of Death: How DIANA's Dual-Use, Commercially Minded, and Process-Oriented Procurement Strategy Will Help Maintain NATO's Technological Edge

Thomas Dallas McSorley—General Counsel, DIANA [thomas.McSorley@diana.nato.int]

Dr Maciej Macenowicz—Chief Counsel, Rapid Adoption Service, DIANA [maciej.macenowicz@diana.nato.int]

Matthew Maddison—Legal Officer, DIANA [matthew.Maddison@diana.nato.int]

Christopher Yukins— Professor, George Washington University Law School [cyukins@law.gwu.edu]

Abstract

The North Atlantic Treaty Organization (NATO) has launched the Defense Innovation Accelerator for the North Atlantic (DIANA) – a unique effort among NATO partners to harness emerging technologies for the Alliance's collective defense. DIANA relies on a network of accelerators and test centers across the NATO Alliance to identify, demonstrate and validate novel solutions, with support from scientists, investors, industry partners, end users and government procurement experts. DIANA will focus on key technologies such as big data, artificial intelligence (AI), autonomy, quantum, biotechnologies and human enhancement, energy and propulsion, novel materials and advanced manufacturing, hypersonics and space, with an emphasis on dual-use (civilian and defense) technologies which can be used to address emerging defense and security challenges. DIANA's board of directors, which is responsible for governance, includes representatives from every NATO country. This paper will explore the special procurement challenges presented by DIANA to ask how best to achieve DIANA's critical goals. The primary area of investigation will be into existing models for defense innovation in the U.S. Defense Department and the European Union, and the research methodology also will look to primary (directives, laws, trade agreements and the acquis communautaire [the accumulated body of European Union law and guidance]) and secondary sources to recommend optimal procurement strategies to meet NATO's unique requirements. The research goal is to facilitate this and future efforts in shared defense innovation to ensure that NATO maintains its technological lead in an increasingly hostile security environment.

Introduction

The recently published *Annual Threat Assessment of the U.S. Intelligence Community* has recognised that

> [n]ew technologies … are being developed and are proliferating at a rate that makes it challenging for companies and governments to shape norms regarding civil liberties, privacy, and ethics. The convergence of these emerging technologies is likely to create breakthroughs, which could lead to the rapid development of asymmetric threats…. [The] competition also exploits technological advancements … to gain stronger sway over worldwide narratives affecting the global geopolitical balance, including influence within it. (Office of the Director of National Intelligence, 2024, pp. 30, 6)

Similarly, the U.S. National Defense Industrial Strategy notes that

> adversarial domination of critical markets allows … control of commodity pricing and access to materials in strategically critical areas…. Much of the civilian manufacturing sector and some of the defense sub-tier supply chain has moved

> offshore into a range of foreign producers, some of whom have become adversarial states … [leading to concern] that predatory adversarial investment and acquisition strategies, often focusing on critical or innovative technologies, further weaken U.S. industrial supply chains and the defense industrial ecosystem's ability to provide capabilities and secure sensitive technologies. (Department of Defense [DoD], 2023, p. 44).

Across the Atlantic, the new European Defence Industrial Strategy highlights that

> [a]dversaries have engaged in a global race for technological supremacy…. Strategic competitors are investing heavily in military capabilities, defence industrial capacities and critical technologies, whilst the integrity of our supply chains … can no longer be taken for granted. … [As such a] technological cutting edge and capacity to steadily guarantee the availability of any defence equipment are prerequisites to the ability of the Union to guarantee the effectiveness of its Member States's armed forces and thereby to preserve peace on the continent. (European Commission, 2024, pp. 2, 3, 31)

In short, there is broad agreement that a critical transatlantic defence and security challenge is the need to access more existing and emerging commercial technology and capacity, more quickly.

In response to this challenge, and to sharpen NATO's technological edge, NATO Allied heads of state and government endorsed the launch of the NATO Defence Innovation Accelerator for the North Atlantic (DIANA). "DIANA will accelerate emerging and disruptive technological solutions – particularly technologies primarily geared for commercial markets that also have potential defence and security applications ('dual-use')" (DIANA, 2022), enabling Allied nations to work "together with the private sector to adopt and integrate new technologies and shape standards" (DIANA, n.d.-a). This article will demonstrate the potential of DIANA to complement existing procurement systems and efficiently provide for the adoption of new, commercial technologies by defense and security end users within the NATO Alliance. While presenting the operational model of DIANA, special attention will be given to the "Rapid Adoption Service" – DIANA's constituent element intended to support agile and rapid development and adoption of innovative solutions by Allies and NATO. DIANA will operationalize the Rapid Adoption Service through a single set of rules supporting a contracting vehicle that can be utilised by multilateral, multinational, and bilateral consortia to continue development and eventually procurement of innovative solutions, directly or through partner NATO elements such as the NATO Support and Procurement Agency.

This article will proceed in eight parts. Strategic Harmonization and Synergy discusses the internationally recognised role of innovation in public procurement, NATO's goals in launching the DIANA initiative to foster innovation in dual-use technologies, and how DIANA's work will be guided strategically. Capability Driven Dual-Use discusses how dual-use technologies will be vetted through DIANA's "Challenge Programmes" to ensure that they meet the needs of NATO and of the broader market. Harmonized and Agile Procurement reviews DIANA's Rapid Adoption Service, which will allow DIANA to facilitate harmonized and agile procurement. Standardization, Interchangeability and Interoperability discusses how DIANA will advance standardization and interoperability, leveraging commercial solutions to improve defense security across NATO. The sixth part, Protected, discusses DIANA's strategies for taking up innovative and strategically important technologies to protect those technologies from potential adversaries. New Production Capacities reviews how the DIANA network of accelerators, test centers, innovators, and investors can advance production among both traditional and non-traditional defense suppliers. The Conclusion concludes with a summary of this brief piece, noting the innovative approaches that DIANA will bring to the development of dual-use technologies in NATO.

Strategic Harmonization and Synergy

According to the OECD's (2017) Public Governance Review, "public procurement is increasingly recognised as a strategic instrument and policy lever for achieving government policy goals, such as innovation …" (p. 16) [but] "must be deployed strategically in coordination with other policy areas" (OECD, 2017, p. 12). Innovation embedded in overarching strategy not only helps to coordinate between different levels of government but also contributes to "coherence and added value in the form of avoiding unnecessary duplication and using synergy effects" (OECD, 2017, p. 52). Furthermore, it sends a strong signal of political commitment that facilitates the process of implementing innovation at every level of administration and governance.

Allied Leaders agreed to launch DIANA in order

> to foster transatlantic cooperation on critical technologies, promote interoperability among Allied forces and harness civilian innovation by engaging with academia and the private sector; … [to] harness the opportunities presented by emerging and disruptive technologies, boosting NATO's competitive edge in collective defence and security; … [and to develop] new capabilities [that] will improve the Alliance's ability to respond to conventional threats – and to the threats posed via these technologies themselves. (NATO, n.d.-a)

These goals have been enshrined in DIANA's Charter, which includes an

> instruction to accelerate civil-military emerging and disruptive technological solutions—particularly dual-use ones—to critical transatlantic defence and security challenges, leveraging existing elements from NATO nations and NATO bodies and guided by relevant NATO Strategies and Frameworks to the effect that high-level objectives for DIANA are based on critical challenges facing operational end users from NATO nations and longer-term strategic priorities that fit within the NATO mission. (DIANA, 2022)

In the wider security context, DIANA can also contribute to enhancing the Alliance's industrial base and its capability to maintain its technological edge, delivering what is needed, when it is needed, without restrictions stemming from excessive external dependencies or bottlenecks. Recognizing that insufficient investments can lead to increased capability and industrial gaps and to greater strategic dependencies, DIANA offers an opportunity to facilitate coordination, pooling defense planning and co-operation in common strategic domains (related to emerging and disruptive technologies) by focusing nations' efforts and resources in the form of targeted investments that will amplify impact to develop and operate the full spectrum of capabilities, "avoiding duplication and increasing efficiency" (European Commission, 2024).

Achievement of such goals begins with the formation of DIANA's 'Strategic Direction.' The Strategic Direction defines strategic-level problem sets and innovation fields for exploration based on critical challenges facing operational end users as well as longer term priorities that fit within the Alliance's political-military objectives. Drawing on inputs from a wide community of operators, scientists and technologists across NATO and Allies, including insights gained from the NATO Industrial Advisory Group and the Conference of National Armaments Directors, and capability gaps identified through the NATO Defence Planning Process (NDPP). The NDPP "provide[s] a framework within which national and Alliance defence planning activities can be harmonised to enable Allies to provide the required forces and capabilities in the most effective way. …[It] aims to… minimise duplication and maximise coherence across the various planning domains" (NATO, n.d.-b). Whereas NDPP documents identify priorities and focus areas both for NATO as a whole and for individual Allies, the Strategic Direction specifies the scope of priorities to be met through DIANA. To that effect, the formation of the Strategic Direction takes into account additional factors such as alignment with defense industrial strategies, promotion of defense industrial cooperation, as well as enhancement of NATO-industry (NATO, n.d.-c). The NATO Science and Technology Strategy also influences the process, as it "enables the generation and exploitation of scientific knowledge and technological innovation as an essential support for the Alliance's core tasks … and inform[s] Allies' investment decision priorities by facilitating the match-making between national and NATO demand" (NATO, 2018, p. 3). The formation of the Strategic Direction places DIANA in the wider context of holistically understood security where inter-reliant sectors (military and non-military) collaborate to deliver protection against both current and future threats of various natures. Through harmonization with this broad range of strategies and policies across the Alliance, DIANA offers a platform to facilitate coordination of planning domains. The intended effect is synergy with respect to the development and adoption of innovative solutions throughout NATO nations.

Capability Driven Dual-Use

DIANA's Strategic Direction is operationalized through DIANA 'Challenge Programmes.' Each Challenge Programme outlines a critical problem statement that can be addressed through innovation fields prioritized in the Strategic Direction (DIANA, 2022). For example, DIANA's 'pilot year' challenges sought solutions in the areas of energy resilience, secure information sharing, and sensing and surveillance. According to DIANA's Charter, "Challenge Programmes are only conducted if there is strong potential for adoption across a number of Allies (or by NATO), if they support the dual-use intent, and if the solutions are likely to be commercially viable" (DIANA, 2022).

The first condition Is achieved by ensuring that challenge formation is aligned to NATO's capability needs identified by the two Strategic Commands (Allied Command Operations and Allied Command Transformation), as well as to national capability needs, that overlap with the Strategic Direction. In effect, DIANA will "focus on the most critical capability gaps and embed with the war fighter to do so" (Beck, 2024, p. 5). The capability-driven approach is intended to support harmonization of procurement and budgetary planning, discussed in the following sections. Furthermore, noting that "uncoordinated outreach, has sometimes resulted in overlapping, unprioritized, and competing demand signals that can make it hard for tech companies to engage, particularly small companies and startups" (Beck, 2024, p. 6), DIANA aims to leverage the potential of capability-oriented challenges as a clear demand signal that reduces risk for emerging technology start-up and scale-up companies, investors, and existing prime contractors as well as more fragile sub-tier suppliers. This can contribute to promoting focused investment into the capacity of the industrial base (DoD, 2023, p. 39).

The second and third conditions (dual-use application and commercial viability) help to address a broader spectrum of threats to security but also contribute to economic security within the Alliance and to building resilient supply chains. Recent developments in the security environment "have uncovered material gaps in the ability of … [the] international [defense industrial base] to rapidly scale production[;] … global supply chains are critical components of … [the] defence industrial ecosystem, yet they are vulnerable, particularly in their sub-tiers" (DoD, 2023, pp. 20–21). "The overreliance on third countries' supplies further undermines security of supply and freedom of action in case of crises" (European Commission, 2024, p. 5), as they may choose to prioritize their own demand over contractual obligations. On the other hand, "[d]eveloping secure alternative sources can involve years-long lead times to reach production scale" (DoD, 2023, pp. 20–21) and therefore requires anticipating action. "Nevertheless, supporting ramp-up also requires dealing with the industrial consequences of a ramp-down once the surge in demand has been met" (European Commission, 2024, p. 17).

DIANA has strong potential to address these issues. First, DIANA is uniquely positioned to pool demand for commercial innovation, complementing existing NATO structures that the conventional industrial base. Second, alongside developing complete systems where appropriate, DIANA will pursue opportunities for the integration of dual-use solutions that do not immediately form a ready-for-use technology into a larger system that may be provided by system integrators. To that end, DIANA will encourage the adoption of open architecture principles and use of, whenever possible, widely accepted industry standards in the design and development of platforms considering that "open architecture allows components to be modular and interchangeable, making it easier to integrate new technologies and updates across different systems" (DoD, 2023, pp. 30–40). This will increase the commercial viability of those solutions by connecting them to customers in the traditional defense sector and, by doing so, will enable utilization of small and medium enterprises' production potential by incorporating into supply chains of critical capability suppliers. Third, leveraging the commercial aspect of emerging technologies with defense applications allows the Alliance to bolster industry's resilience against cyclic demand and shifts in defense budgets which have traditionally hampered the ability of sub-tier suppliers to remain in the defense market. DIANA solution providers will be able to maintain their production lines to supply their civilian customer base while remaining available to defense and security applications. Thus, it is critical that DIANA solutions are developed on a true "dual-use" path for as long as possible. That is, these solutions will not be "forked" (splitting the "defense" and "civilian" versions) for as long as possible to avoid fragmenting their development cycles and their production lines. This lowers the costs of development and makes for a better product, where the base version is the sum of the best and most talented development team and process without regard to end use. This is accomplished in part by bringing a commercial perspective to the DIANA solution pipeline. Fourth, the holistic approach to security and understanding of interdependencies between different sectors (especially in regard to critical infrastructure) will allow focus on those dual-use technologies that are in demand in times of peace, crisis and conflict, using broad market fit to mitigate the risks of failure that befall early-stage and deep tech companies. Finally, maintaining the dual-use aspect of DIANA solutions will facilitate the scaling up of production, leading to reduced costs and increased availability, or as the EU describes, it contributes "to the building up of 'ever-warm' spare industrial capacities that allow for the necessary flexibility to ramp up in response to urgent spikes in demand" (European Commission, 2024, p. 17).

Each DIANA Challenge Programme features a competitive selection process—organised by DIANA—with bids submitted by innovators from across the Alliance. To avoid the hampering effect that prescriptive requirements inflict on innovation (OECD, 2017, p. 52), the functional approach is utilized by DIANA in pursuit of the best solution to the identified needs. The selected innovators enter a pipeline where DIANA leverages its affiliated 'Accelerator Sites' and 'Test Centres' to iteratively demonstrate each innovator's proposed solutions to relevant end user operators across the Alliance to receive feedback and determine 'product-market' fit. Vetted innovators with the most promising solutions will be able to continue their work utilizing DIANA's affiliated and constituent elements to make further progress with their proposed solutions to the specific challenges. Continuous dialogue with operational end users, capability developers, system integrators, industrial partners, and investors will ensure the technological solutions fostered within DIANA continue to address the identified problem sets for Allies and NATO, maintaining their commercial viability. Innovators' journeys through DIANA's pipeline will take place in successive stages, where discussion of all relevant aspects with the innovators will be ongoing and iterative. These discussions will include, but will not be limited to, technical aspects, economic aspects (prices, costs, revenues, etc.) and legal aspects (distribution and limitation of risks, guarantees, possible creation of special purpose vehicles, etc.). Both potential for adoption within the defense and security marketplace and overall commercial viability will serve as guiding principles of the above-mentioned dialogue, with a possibility to discontinue any solution within the Challenge Programme at each stage.

[1] Selected innovators are successful participants in the challenge pipeline and will be eligible to continue iteration, development, and production within the DIANA ecosystem through the Rapid Adoption Service without further competition.

Harmonised and Agile Procurement

DIANA's Rapid Adoption Service aims to be a quick, efficient and opt-in vehicle based on a single, agile set of rules that provide:

- support to innovators in the transition and delivery of interoperable solutions to relevant market(s), including navigating Allied and NATO procurement opportunities;
- agile contracting for emerging and disruptive solutions through multinational, multilateral, and bilateral programs that allow for the procurer's challenge and the innovator's solution to be continuously modified with performance-based milestones;
- agile contracting methodologies allowing Allies and/or NATO to leverage DIANA as a complement to their procurement system, pooling demand and simplifying processes where appropriate.

Operation of the Rapid Adoption Service will be supported by a dedicated digital platform collecting, processing, matching, and making available information on supply and demand to stakeholders as well as facilitating process management. The Rapid Adoption Service will allow operational end users to follow the development of dual-use solutions in the DIANA ecosystem to determine their interest in entering follow-on contracts with innovators and, depending on the case, with system integrators and investors as part of a consortium. The scope and the objective of the resulting legal arrangement will be determined on a case-by-case basis, but such arrangement will be based on a modular platform that does not require the typical lead times and complexities of one-off multilateral efforts.

The scope of agreements provided as a part of the Rapid Adoption Service will include but not be limited to:

- **Research and Development Agreement –** concluded to identify technical and performance characteristics of a new product that will satisfy the capability needs of an end user, to confirm the achievement of the level of technological readiness and to develop solutions further, as deemed necessary by participating nations, including development of a *prototype* showing the application of a new concept or new technology in real life or representative environments.
- **Research and Development Agreement with Optional Procurement –** applicable to cases whereupon positive verification of the prototype the participants to the agreement will agree on transition to a *manufacturing subphase,* cooperating on testing, verification and certification as needed to begin delivery of the final product to the end user.
- **Procurement Contract** – used for the acquisition, licensing, purchase, or employment of a complete product and its adaptation to specific technical requirements of an end user (if needed).

In executing its adoption function, DIANA will closely cooperate with other NATO bodies such as the NATO Support and Procurement Agency and the NATO Communications and Information

[1] See: European Commission (n.d.).

Agency to leverage existing resources and expertise. This structure will enable the Rapid Adoption Service to provide all necessary procedural support and act as a broker and/or agent on behalf of those Allies that have limitations in their ability to contract with 'foreign' (Allied) innovators.

The development of the single set of rules supporting the Rapid Adoption Service is guided by the principle that "the harmonization of laws aims to transform the public procurement system into more strategic system in which promotion of innovation plays a central role" (Gomes, 2024, p. 204). DIANA notes that despite substantial efforts to enhance common rules governing European innovation procurement, "harmonisation of national laws (still) faces considerable difficulties where it clashes with administrative law traditions" (Gomes, 2024, p. 105), as different nations may understand and implement common principles differently. Furthermore, as the OECD observed, government "agencies responsible for public procurement are often not co-coordinated with agencies and ministries in charge of innovation policies" (OECD, 2017, p. 17), while to "promote innovation, contracting authorities need discretion and flexibility in both the procedural and the contract execution phases" (Gomes, 2024, p. 201). To address these difficulties, DIANA will leverage available expertise by engaging with experts and seeking continued input from national authorities and international organizations to fully understand the specifics of their legal and procurement system. DIANA aims to connect and complement existing systems, with a new harmonized and coordinated instrument covering all essential stages of innovation procurement to amplify best practices developed by DIANA and national innovation entities across the Alliance's innovation ecosystems and facilitate their equal implementation through the single, agile set of rules, contracting methodologies and supportive management processes which will enable quick and agile procurement.

The potential of the Rapid Adoption Service should also be considered in the context of strengthening the defense industrial base due to consolidation of demand and supply through multinational, multilateral and bilateral programs enabling collaborative procurement. As discussed in previous sections,

> demand is still largely organised along national lines, with most investment decisions arising from domestic considerations, and based on national programming, often failing to factor in broader strategic and efficiency considerations. As a result, … the supply side remains also essentially organised along national lines… This results in a scattered [defense technological and industrial base], … [and in] duplications and foregone opportunities. (European Commission, 2024, p. 5)

In response to this issue, the Rapid Adoption Service will offer a "networked cooperative framework" for Allies that enhances defense industrial output, seeks to de-risk supply chains and improve resilience by growing multiple "production lines across a consortium of like-minded nations" (DoD, 2023, pp. 22, 23). Correspondingly, the alignment with strategies and policies, and the capability drive for co-development with operational end users is designed to create conditions leading to collective demand that can be channelled and subsequently fulfilled through multinational programs that create critical mass. This aims to incentivize the supply side to cooperate and seize economies of scale, allowing quicker production and procurement, at lower cost, boosting defense production, innovation, and overall capability.

Leveraging the multinational/multilateral formula combined with DIANA's current network of over 180 Test Centres and access to NATO's resources of operational experimentation and testing, the Rapid Adoption Service will additionally help to avoid the rigidity of typical linear procurement, combining the aforementioned with the potential of business development opportunities provided by over 20 Accelerator Sites and strengthening emphasis on interoperability, interchangeability and standardization, as explained in the next section.

Standardization, Interchangeability and Interoperability

Standardization contributes to interoperability of Allied forces as well as to reinforcement of the respective industrial bases and increase in overall production capacity. Unfortunately,

> [d]espite the Standardisation Agreements (STANAGs) adopted in the framework of NATO, the voluntary uptake of these standards remains an issue. … [C]urrently agreed standards often do not sufficiently deliver the requisite real interoperability and interchangeability in operational terms since they do not cover all defence systems, nor are they systematically detailed enough. … [C]ertification also remains an issue … which de facto fragments the market and hampers logistics. (European Commission, 2024, pp. 10–11)

DIANA's focus on new (innovative) dual-use solutions developed through the international formula (presented above) in engagement with end users, system integrators, and investors offers an opportunity to facilitate the process of standardization avoiding, at the same time, the risk of hampering innovation that may be caused by setting up requirements and standards too early in the process. To that end, DIANA will follow the OECD's recommendations for using a functional approach at the beginning of DIANA's Challenge Programmes (OECD, 2017, p. 52) and then allowing the "procurer's challenge and the innovator's solution to be continuously 'tweaked' with performance-based milestones" (DIANA, 2022). Co-development, alongside operational end users and agile capability developers from the early stages will enable assessment of the technology in an operational environment leading to alignment of product development with end-user needs and industrial roadmaps. Furthermore, it will provide an opportunity to reach an agreement on applicable standards, or on the development of new standards, as well as on mutual recognition of certification, accreditation, and test results, which can be facilitated through enhanced cooperation within DIANA's network of Test Centres acting as a central node for engagement between innovators and operational end users.

In this process, DIANA will leverage its dual-use focus and utilize "widely accepted industrial standards" rather than operate in isolation or invent new, narrowly applicable standards, which "will facilitate and simplify integration and production efforts" (DoD, 2023, p. 35). This is because "increasing standardization allows for economies of scale and streamlined production processes, and greater interoperability"; and can … help "small businesses and non-traditional suppliers work with [defence]"; further, [t]hey reduce barriers to entry by simplifying product development and integration, making it functionally easier and less expensive for these suppliers to participate and compete, and focus on niche areas of expertise and contributing innovations to the broader defence ecosystem" (DoD, 2023, pp. 35, 36).

Protected

"Ensuring sufficient access to finance for … the defence sector is vital given the compelling need to boost investment in this ecosystem," and given "[small and medium enterprises] operating within the defence sector face higher barriers to accessing finance compared to companies active in other sectors" (European Commission, 2024, p. 24). It has been noted that "limited funding or financial related problems appeared to be the main cause driving startups down the Valley of Death curve" (Gbadegeshin et al., 2022, p. 4). In absence of sufficient funding, small cutting-edge companies "are too often bought up by larger overseas companies before they can develop into the medium sized enterprises" (House of Commons: Science and Technology Committee, 2013, p. 3). This creates opportunities for "adversarial nations that are strategically employing investments in key U.S. and allied defence industries to harvest critical technologies, gain access to pioneering innovation and research and development efforts, … and capitalize on dual-use technologies that may be used to close the gap in… [NATO's] comparative advantage." Failure to "protect … critical industries and assets from adversarial influence … could lead to strained diplomatic relations, decreased trust, loss of foreign defence sales to competitors (perhaps including adversaries), and possibly even weakened economic ties, rendering deterrence of aggressive behaviour by adversarial nations more difficult" (DoD, 2023, p. 49).

In response to the above problem, technologies critical for NATO defense and security challenges in the DIANA ecosystem are incubated in a transatlantic network running an accelerator program which makes innovators cognisant of the particularities of the defense and security sector and of best practices to protect their technologies from nefarious actors. This effort is complemented by the 'Allied Capital Community,' a pooling of capital resources which helps deny nefarious actors free rein in targeting and transferring critical technologies. It aims to achieve this through setting a common and consistent baseline, delivered through a digital platform where Allies can share and exchange the results of vetting procedures on capital investors (capital supply) and relevant innovators (capital demand), including those related to ownership and control. Through collaboration, Allies and the private sector can benefit from increased transparency in this area and, at national discretion, potentially avoid the duplication of vetting efforts. Acknowledging that "private investment increases as companies move from concept to prototype to product" (Defense Innovation Board, 2023, p. 4), the Rapid Adoption Service will allow investors, associated with the Allied Capital Community, to follow technological developments of DIANA's solutions and to participate in dialogues between innovators and end users. Recognizing that "[f]inancial actors' willingness to engage with the defence industry appears to be affected by specificities of the defence market (including complexity of procurement)" (European Commission, 2024, p. 25), the availability of streamlined and consistent procurement processes described above are designed as the economic incentive for investors to financially engage in the development of cutting-edge technologies, with commercial viability and security applications, that require capital support. This contributes to the economic security and technological edge of the Alliance.

New Production Capacities

Reaching out to non-traditional defense suppliers lies at the core of DIANA's mission, so significant effort must be made to "build and deepen relationships with commercial industries not traditionally involved in defence work" (DoD, 2023, p. 20). For example, in the United States, "federal contracting to small businesses owned by underrepresented socio-economic groups accounts for less than 10% of all federal contracting dollars. These suppliers come from diverse industries and can bring technological, production, and process advancements to the defence sector" (DoD, 2023, pp. 19–20), broadening the industrial base and fostering competition within the market. Unfortunately, "high barriers to entry disincentivize the types of small or sub-tier suppliers that help to diversify and make the industrial base more resilient" (DoD, 2023, p. 20). Eliminating unnecessary bureaucracy and encouraging industrial cooperation would help make possible the quick delivery of critical defense components from each nation's respective industrial bases and has therefore been advocated as a measure to improve security of supply both for the United States (DoD, 2023, pp. 21–22) and the European Union (European Commission, 2024). The Rapid Adoption Service, with its agile contracting methodologies and tools for navigating procurement activities, to contribute to that endeavor while the holistic approach to security reflected in the dual-use emphasis of DIANA will support the integration of innovative solutions into multiple and diverse supply chains, including that of Allied system integrators. Recognizing the importance of leveraging private market resources and knowledge and enabling communication in a collaborative manner, as well as promoting cross sector discussions (Kuchina-Musina & McMartin, 2024, p. 70), DIANA has designed its accelerator program to equip businesses with the skills and knowledge to navigate the world of deep tech, dual-use innovation and the defense marketplace through a combination of lectures, workshops, and mentorship that will help participants to build their companies into viable dual-use ventures shaping a peaceful future for the Alliance (DIANA, n.d.-b).

Conclusion

1. DIANA is a new NATO body established to help the Alliance build and keep its strategic advantage in the field of emerging and disruptive technologies.
2. By harmonization with the broad range of strategies and policies that inform strategic planning, DIANA can leverage coordination of planning domains contributing to synergies and effectiveness of innovation efforts within the Alliance.
3. Thanks to its capability-driven dual-use approach, DIANA will foster technological solutions to address military needs of operational end users, maintaining their commercial viability with a view to increasing overall industrial production capacity.
4. DIANA's Rapid Adoption Service, based on a single set of rules, will allow for quick and agile international collaborative procurement utilizing collective demand signals to achieve effects of scale and incentivize investment on the supply side.
5. DIANA's focus on innovative dual-use solutions developed in the international forum in collaboration with end users, system integrators and investors will offer an opportunity to facilitate the process of standardization, foster interoperability and strengthen cooperation within the Alliance's industrial base.
6. DIANA provides a mechanism designed to support start-ups through all stages of development; protecting them against adversarial and/or existential financial, security, and compliance risks; and enabling them to build their solutions into viable dual-use ventures.
7. Strengthening communication and engagement with industry will support NATO in accessing the potential of start-ups and small and medium enterprises to grow resilient supply chains of critical components and technologies across the Alliance.

References

Beck, D. A. (2024). *DIU 3.0: Scaling defense innovation for strategic impact*.

Defense Innovation Board. (2023). *Terraforming the valley of death - making the defense market navigable for startups*.

DoD. (2023). *National defense industrial strategy*.

DIANA. (n.d.-a). https://diana.nato.int/about-diana.html

DIANA. (n.d.-b). https://diana.nato.int/accelerator-programme.html

DIANA. (2022). *Charter of the NATO defence innovation accelerator for the North Atlantic*.

European Commission. (n.d.). *Explanatory note – competitive dialogue – classic directive*.

European Commission. (2024). *Joint communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A new European defence industrial strategy: Achieving EU readiness through a responsive and resilient European defence industry*.

Gbadegeshin, S.A. et al. (2022). Overcoming the valley of death: A new model for high technology startups.

Gomes, P. C. (2024). *EU public procurement and innovation: The innovation partnership procedure and harmonization challenges*.

House of Commons: Science and Technology Committee. (2013). *Bridging the valley of death: Improving the commercialisation of research*.

Kuchina-Musina, D., & McMartin, B. (2024). *Public procurement for innovation*.

NATO. (n.d.-a). https://www.nato.int/cps/en/natohq/topics_216199.htm

NATO. (n.d.-b). https://www.nato.int/cps/en/natohq/topics_49202.htm

NATO. (n.d.-c). https://www.nato.int/cps/en/natolive/topics_49160.htm

NATO. (2018). *NATO science and technology strategy*.

OECD. (2017). *Public procurement for innovation: Good practices and strategies*.

Office of the Director of National Intelligence. (2024). *Annual threat assessment of the U.S. intelligence community*.

Acquiring Technology for Allies and Partners

Matthew Reed—is a researcher at the Institute for Defense Analyses (IDA). [mreed@ida.org]

Abu Naimzadeh—is a researcher at the Institute for Defense Analyses (IDA). [anaimzad@ida.org]

Jarrett Lane—is a researcher at the Institute for Defense Analyses (IDA). [jlane@ida.org]

EunRae Oh—is a researcher at the Institute for Defense Analyses (IDA). [eoh@ida.org]

Jennifer Taylor—is a research staff member at the Institute for Defense Analyses (IDA). [jtaylor@ida.org]

Abstract

The Institute for Defense Analyses (IDA) produced and submitted this paper for the Naval Postgraduate School's 21st Annual Acquisition Research Symposium, May 8–9, 2024, in Monterey, CA, as a discussion draft.

The Defense Security Cooperation Agency (DSCA) asked IDA to assist in developing new approaches to providing capabilities to partner nations—particularly commercially available capabilities that could be used by partner nations to exercise self-defense and deterrence vis-à-vis the People's Republic of China (PRC) and Russia. This paper offers a summary of research, findings, and potential recommendations for how the Department of Defense's (DoD's) security cooperation community, DSCA, and defense innovation organizations (DIOs) can develop new business strategies for incorporating commercially available capabilities into security assistance activities.

The project team's methodology included expert interviews and literature reviews. The project team conducted structured interviews with government, industry, and IDA experts on commercial technologies and their potential military applications, defense innovation initiatives, and security cooperation programs. The project team completed literature reviews to understand known and potential uses of commercial technologies by partners nation militaries. Finally, through a combined approach of expert interviews and literature reviews, the project team developed a heatmap designed to highlight which commercial technologies may yield the greatest impact on the ability of a partner nation's military to execute critical tasks. It is the IDA team's hope that this paper will elicit constructive feedback and insights that can further shape our research and recommendations to DSCA.

The 2022 U.S. National Defense Strategy (Austin, 2022) highlights the importance of leveraging commercial technologies and innovation to maintain U.S. warfighting advantages. Further, the use of commercial technologies in the Ukraine war demonstrates their utility and impact in modern conflict. While the DoD is increasingly investing in commercial technologies, there is no strategy for helping partner nations understand potential uses for commercial technologies; identify viable solutions in the commercial market; and acquire, integrate, and use commercial technologies in warfighting and task execution. DSCA's strategy and processes for enhancing partner nations' capabilities lean heavily on sales and transfers of hardware, platforms, and systems that were developed for the U.S. military and covered by current and past programs of record. These are often more expensive, harder to maintain, and more difficult to replace than commercially available capabilities.

Problem Statement

Commercial technologies can improve partner nations' capabilities across functional areas and tasks critical to self-defense and deterrence. However, the Defense Security Cooperation Agency's (DSCA's) current approach to enhancing partner nations' capabilities leans heavily on sales and transfers of traditional hardware, platforms, and systems—both legacy items and products covered by existing programs of record—developed for the U.S. military. This approach enables interoperability with U.S. forces, ensures sustainability and long-term supportability, and promotes the U.S. defense industrial base. However, these systems are often expensive and difficult to maintain, and they may take years to acquire and deliver. Further, many of these capabilities were developed in the context of U.S. military missions and the way in which the U.S. warfighter will act in conflict, which may be different from the approach of a partner military. Helping partner nations incorporate commercial technologies into their operations presents an opportunity to both improve their ability to counter Chinese and Russian aggression and improve the U.S.'s position as a partner of choice.

In an effort to displace the United States and establish itself as the partner of choice, the People's Republic of China (PRC) is capitalizing on shortcomings in the United States' current approach to security assistance (e.g., expensive items and long acquisition cycles) by incorporating cost-effective and agile commercial, dual-use technologies into its security assistance packages and foreign military sales. For example, Huawei sold cybersecurity capabilities to Indonesia's National Cyber and Crypto Agency—roughly equivalent to the U.S. National Security Agency (Syamsudin, 2023). The PRC can often provide its commercially available capabilities faster and cheaper than the United States, creating technological and economic dependencies that provide the PRC with global influence among U.S. partners (Nagar, 2022; Nouwens & Legarda, 2018; Russel & Berger, 2020).

Over the last several decades, the PRC drastically strengthened its commercial and dual-use technology industries. The PRC's civil-military fusion strategy, coupled with an increase in the country's defense spending, led to tens of billions of dollars in funding to streamline PRC commercial research and development efforts and to advance dual-use technology projects (Cheng, 2023; China Power Team, 2018; DoD, 2023; Nouwens & Legarda, 2018). The Chinese Communist Party even created the Central Commission for Integrated Military and Civilian Development to promote collaboration among Chinese universities, technology companies, and the military, making it difficult to bifurcate Chinese commercial activity from government and military activity (Nouwens & Legarda, 2018). Between 2022 and 2023, the U.S. Department of Commerce added more than 70 Chinese technology companies to the Entity List[1] after determining the companies have "close ties . . . to the Chinese military and the defense industry" and that they often specialize in dual-use technologies like artificial intelligence (AI) and software used for weapon life cycle management (Additions and Revisions to the Entity List and Conforming Removal from the Unverified List, 2022, p. 3; Additions of Entities to the Entity List and Removal of Entity from the Entity List, 2023). In fact, according to a 2022 RAND report, almost half of the PRC's manufacturing output is considered dual-use (Weinbaum et al., 2022, p. 4).

The PRC's civil-military fusion strategy also aligns with its Belt and Road Initiative—a strategy for growing the PRC's economic and political global influence (Syamsudin, 2023). Many export-restricted companies on the Entity List, like ZTE, Huawei, Tencent, Hikvision, Zhejiang Dahua, and Alibaba, have directly sold their dual-use technologies (e.g., cloud, AI, and surveillance technology) to foreign governments, militaries, and security forces on almost every continent (Bouey et al., 2023; Montgomery & Sayers, 2023; Sahin, 2020). Since 2009, Chinese technology companies have sold cyber capabilities to Indonesia's National Cyber and Crypto Agency and signed contracts with more than 140 countries to install automated "safe city" surveillance equipment (including facial recognition technology) that allows governments to monitor cities and towns (Kynge et al., 2021).

The PRC can also lean on ubiquitous availability of some Chinese-produced technologies—including commercial drones—to foster partnerships and become a source of

[1] The Entity List is a U.S. government list of foreign individuals, companies, and organizations deemed a national security concern, subjecting them to export restrictions and licensing requirements for certain technologies and goods.

low-cost, quickly acquirable capabilities. DJI, a Chinese commercial drone company, accounts for 70% of the global drone market (Anwar, 2023). In 2023, the Chinese government imposed new exports controls on commercial drones, ostensibly to curb militarized use of DJI and other Chinese-produced commercial drones (McDonald, 2023). However, DJI and other Chinese drones remain readily accessible and used for military purposes (Gosselin-Malo, 2023). Thus, helping partner nations acquire and use U.S. and/or allied-produced commercial technologies is important to ensuring the United States remains the partner of choice for national defense capability and capacity building.

Leveraging commercial technology is not only essential to defending American alliances and influence abroad, but is also key to helping U.S. partner nations deter and defend themselves against aggression in cost-effective and adaptable ways. The war in Ukraine highlights the potential impact of helping partners apply commercial technologies to bolstering deterrence and self-defense vis-à-vis larger aggressors. The first salvo in Russia's invasion was massive, widespread cyberattacks. In response, the Ukrainian government rapidly transitioned data and operations to commercial cloud environments which significantly improved the government's resilience against cyberattacks. Using edge computing devices, the Ukrainian government worked with U.S. cloud companies to move terabytes of data to commercial cloud infrastructure at the beginning of Russia's invasion. This decision enabled the Ukrainian government to preserve critical data and government services and improve its resiliency against continued cyberattacks. At an industry event in November 2022, Ukraine's Vice Prime Minister and Minister of Digital Transformation, Mykhailo Federov, explained,

> We experience cyberattacks on a daily basis; this is a cyberwar. Under attack is our critical infrastructure. But we have been successful in protecting [it], and every week we are launching some new public resources. Digitalization is the best response to this challenge. (Konkel, 2022)

In the battlefield, commercial technologies are playing a pivotal role in Ukraine's ability to damage Russian forces. Ukrainian warfighters are innovating with commercial drones to improve reconnaissance, targeting, and delivery of small munitions. Agile software development practices and tools used in commercial sectors enabled the Ukrainian army to develop Delta, a software platform that integrates satellite imagery, drone imagery, social media, and more to create a comprehensive view of the battlefield and derive actionable intelligence (Borger, 2022). Commercial satellite communications, provided by Starlink, enabled Ukrainian forces to deploy Delta directly into the hands of warfighters, and maintain command and control (C2) after Russia disrupted Ukraine's military satellite communications (Jones et al., 2023).

Ukraine's use of commercial drones also serves as an example of the potential cost and speed advantages of commercial, dual-use technologies. Drones are enabling Ukrainian forces to efficiently and effectively deliver results that are traditionally achieved—at least in the context of U.S. doctrine—through artillery, missiles, and sophisticated platforms (e.g., aircraft). For example, Ukraine is using commercial drones to develop loitering munitions that target armor, weapons systems, entrenched combatants, and more. Though commercially available capabilities may not always be a suitable replacement for traditional capabilities, they can be additive and may effectively augment how partner nations execute critical tasks.

Creative applications of commercially available capabilities cannot replace or eliminate the need for traditional materiel like artillery shells, but they may be useful in mitigating risks in supply shortages and augmenting partners' warfighting plans, and tactics. Ukraine's use of commercially available capabilities are not only cost-effective, they are borne out of necessity. Bottlenecks and capacity constraints in the supply chain for materiel necessitate innovation. As of 2023, Ukraine's monthly use of 155-millimeter shells outstripped one year of U.S. production capacity (Morris, 2023). Though increasing production capacity and shoring-up the defense industrial base is a top priority, challenges such as hiring, raw material availability, and manufacturing equipment will remain a concern for the foreseeable future (Morris, 2023).

The *2022 National Defense Strategy* (*NDS*) emphasizes the importance of innovation and commercial technologies to ensure the U.S. warfighting advantage—a point of emphasis that the *NDS* extends to allies and partners as well (Austin, 2022, p. 19). The strategy states, "The [Department of Defense] will support the innovation ecosystem both at home and in expanded partnerships with our Allies and partners." The *NDS* also notes that the Department will "assist Allies and partners" in improving their resilience and ability to "fight through disruption" by improving, for example, cyber resilience through technologies such as modern encryption and zero-trust architectures. (Austin, 2022, p. 8).

Notwithstanding recognition of commercial technologies' potential impact on partner nations' warfighting capabilities, there is currently no established lead, strategy, or initiative to help partner nations understand potential uses for commercial technologies; identify viable solutions in the commercial market; and acquire, integrate, and use commercial technologies in warfighting and task execution. To help address this gap, the DSCA aims to develop new approaches for providing innovative commercial capabilities to partners.

Defense Innovation Organizations: Insights and Lessons Learned

As a first step to help DSCA identify commercially available capabilities, the Institute for Defense Analyses (IDA) team researched whether defense innovation organizations (DIOs) are incorporating partner nations into their activities, as well as how DSCA could collaborate with DIOs to help partner nations procure and integrate commercially available capabilities. IDA engaged with 17 DIOs and interagency organizations to determine if DSCA can access their data and insights into commercially available capabilities. IDA also captured lessons learned about how the organizations interact with industry and explored opportunities and challenges to leveraging the organizations' expertise and resources to support moderately capable partner nations.[2]

Through this research, IDA found that the DIOs generally lack the mandate, resources, presence, and expertise required to help address partner nations' needs proactively and consistently. DIOs focus foremost on acquiring technologies for the Department of Defense. In cases where DIOs are engaging internationally, it is typically with allies and high-end partners and focused on international armaments cooperation in service of better capabilities for the U.S. warfighter. There are periodic engagements with moderately capable partners, though typically in response to an active conflict or acute problem (e.g., the war in Ukraine).[3] However, the DIOs do not have focused lines of effort or goals associated with helping partner nations leverage commercially available capabilities.

IDA found that individuals throughout DIOs recognize the importance of supporting partner nations and expressed willingness to help DSCA on a case-by-case basis. Working with willing leaders and individuals within DIOs may allow DSCA to tap the DIOs' networks, expertise, and capabilities for identifying, vetting, and sourcing commercially available capabilities. Connecting DIOs' knowledge of commercial capabilities with the expertise of DSCA and Security Cooperation Organizations (SCOs) has the potential to deliver value to partner nations. But, until DIOs are appropriately resourced and directed to support partner nations,

[2] Loosely defined, "moderately capable partner nations" are partners with limited resources that have the potential to fight alongside or in lieu of U.S. forces. These partners are DSCA's priority focus for developing new approaches to delivering commercially available capabilities.
[3] See, for example, the Defense Innovation Unit's support to requirements generation and technology delivery for Ukraine. The DIU indicates it plans to increase its activity with international allies and partners as part of DIU 3.0—the organization's latest evolution—but is prioritizing engagement with allies and well-resourced partners like India (Beck, 2024).

scaling collaboration between DSCA and DIOs will be difficult and engagements will be reliant on the goodwill and bandwidth of DIO team members.

Innovation Organizations: Data Sources

The Undersecretary for Research and Engineering identified 271 DoD organizations as "innovation organizations."[4] One of these organizations, the Defense Innovation Unit (DIU), reported in FY 2022 that they received over 1,600 pitches from industry. That same year, the DIU delivered 82 prototypes and transitioned 17 capabilities (DIU, 2023). These figures suggest DIOs have the reach and market visibility that could help DSCA accelerate and scale efforts to find commercially available capabilities relevant to partner nations. The DIU and other DIOs curate their ecosystems of industry partners and gain visibility into the market, in part, by conducting a variety of industry outreach activities. Examples include open "office hours," pitch events, and formal requests for proposal (RFPs) aimed at attracting non-traditional vendors and innovative solutions.

DIOs and interagency organizations often work closely with trade associations. Working with trade associations (e.g., the Association for Uncrewed Vehicle Systems International) could help DSCA scale outreach and engagement with industry, as well as access datasets maintained by the associations on members and their respective products (AUVSI, n.d.). For example, the Association of Drone Manufacturers maintains a dataset that is accessible for a fee and updated daily with technical specifications and contact information for specific drones (AUVSI, 2022).

Finally, open-source researchers and reports can serve as a valuable source of intelligence on technology trends and applications. For example, some researchers spend significant time and effort tracking the models of drones and other technologies used in active conflicts (e.g., Ukraine).[5] Although not all of these technologies are going to be U.S.- or ally-provided, DSCA can use open-source reporting and researcher datasets to find examples of technologies and products used by military and security forces globally. Additionally, subscription services to specialized publications (e.g., *Jane's*), industry analyst reports (e.g., Gartner), and market intelligence services (e.g., Crunchbase and Futurepedia.io) will give DSCA insights into key trends and potential solutions for partner nations.

Innovation Organizations: Opportunities for Collaboration

IDA met with innovation organizations that recognize the importance and potential impact of helping partner nations acquire commercially available capabilities. In fact, the DIU 3.0 plan states, "We must connect the solutions created by U.S. tech companies to allied and partner acquisition organizations when appropriate . . . especially in a conflict, when speed is critical." Some DIO representatives, including DIU, offered to query their datasets for targeted requests from DSCA and help DSCA run outreach events in support of specific partner nations.

Collaborating with innovation organizations can also help DSCA identify vetted, proven capabilities critical to ensuring partner nations acquire effective solutions and mitigating risks of deploying untrustworthy or unproven technologies. For example, the Department of Homeland Security's National Urban Security Technology Laboratory conducts market surveys and technological testing to inform local emergency response departments of the capabilities available in the commercial market. Though these are more focused on lower-end security

---

[4] Not all of these organizations conduct substantial commercial outreach for mature technologies; others engage in deeper, longer-term technology development, technology assessments, or some other role in the tech development process (DoD Innovation Pathways, n.d.).

[5] See, for example, a publicly accessible spreadsheet covering drone incidents in Ukraine accessible on Google (*Ukraine Drone War Incidents 2024*). This database was highlighted by *Foreign Policy* in a February 2023 article, "The Drone War in Ukraine is Deadly, Cheap, and Made in China" (Greenwood, 2023).

forces, some of these capabilities could have applications for U.S. allies (e.g., unmanned aerial systems [UAS] for first responders, counter-UAS, counter-improvised explosive device, and protection against lasers; U.S. Department of Homeland Security, n.d.-a, n.d.-b; National Urban Security Technology Laboratory, 2022) DHS indicates that they are willing to provide access to non-public documents for other federal agencies (U.S. Department of Homeland Security, n.d.-c).

In addition, some technology accelerators partner with DoD research organizations to technically vet members of their defense portfolios (Director of Private Sector Technology Accelerator, personal communication [phone interview], November 16, 2024). Because these organizations can have hundreds of companies in their programs, DSCA could identify vetted technologies more quickly by engaging with accelerator business development managers for lists of their programs.

The Department of Commerce offered to contact trade associations to discuss technology solutions on DSCA's behalf. However, due to staffing limitations and competing priorities, commerce-driven outreach would likely need to be planned far in advance, limited in scope and frequency, and in response to a clear demand from partner nations. The Department of Commerce regularly hosts webinars and in-person events with partner governments and U.S. industry to help connect U.S. corporations to active RFPs from partner governments,[6] but Commerce does not collect data on U.S. corporate participation, nor does Commerce conduct technical vetting.

Innovation Organizations: Barriers and Challenges to Collaboration

Though DIOs indicated openness to partnering with DSCA, there are legal and policy roadblocks to data sharing. For example, contractual and policy restrictions designed to safeguard proprietary and competitive industry data inhibits the transfer of technical data to other DoD organizations, creating a hesitance or inability for DIOs to provide unlimited data access to DSCA. DIOs can instead query data sources in response to specific requests from DSCA and respond with information on capabilities that may be of interest.

Relying on DIO staffs to respond to data query requests from DSCA presents limitations. Because organizations are limited in staffing and bandwidth, they may be unable to respond to individual or ad hoc requests for data, and thus there is a limit to the pace and scale of requests. This can be mitigated by creating more targeted, specific requests to individual program managers for data already organized. However, scaling data requests is a challenge as the sheer number of DIOs may necessitate dozens, if not hundreds, of requests for DSCA to identify viable solutions for partner nations. It is possible that DIU could provide a pathway for streamlining data requests as DIU takes on responsibility for coordinating activities across DIOs. DIU Director Doug Beck recently wrote,

> Going forward, DIU will work with partners across the Department's community of defense innovation entities—as well as with the Chief Data and Artificial Intelligence Officer (CDAO)—to take advantage of opportunities to generate impact through shared best practices, talent management, shared systems and processes, and enhanced teamwork. DIU has been charged by the Secretary and Deputy Secretary of Defense with ensuring maximum synergy—and eliminating dyssynergy—across this team. (Beck, 2024)

Focusing data collection efforts exclusively on commercial entities who have previously worked with DoD organizations narrows the aperture and potentially overlooks more optimal

[6] See, for example, a March 2024 webinar organized by the U.S. Department of Commerce regarding defense export opportunities in the Philippines (U.S. Department of Commerce Industry & Analysis-Aerospace Office and U.S. Commercial Service, 2024).

solutions for partner nations. Furthermore, while there is substantially lower technical risk by focusing on DoD-vetted commercial technology, it is possible that a higher adherence to stringent DoD requirements leads to a higher cost and complexity than could be achieved from buying off-the-shelf technology. In addition, many DIOs are focused on developing technologies that may not result in prototypes (much less exportable capabilities) in the short term, thus limiting immediate relevance to partners.

Heatmap: Prioritizing Technologies and Capabilities

Following its findings that DIOs are willing to assist DSCA but lack the mission and capacity to scale activities supporting partner nations, the IDA team recalibrated its efforts to helping DSCA determine ways it could lead, scale, and sustain efforts to help partner nations harness commercially available capabilities.

Because DSCA focuses on sales and transfers of traditional materiel and systems today, the IDA team determined that a first critical step is providing DSCA a comprehensive view of commercially available capabilities that may be relevant to moderately capable partners. However, commercially available capabilities will not be appropriate for all missions and tasks, and some capabilities may not be relevant to moderately capable partners. Nor is it feasible for DSCA to pay attention to all commercially available capabilities.

To help DSCA focus on commercially available capabilities that may be most impactful for partner nations, the IDA team developed a heatmap that shows the interplay between technologies and critical tasks (e.g., targeting, joint fires, etc.; see Appendix A). A heatmap is a data visualization method that offers a simple visual representation of the value or magnitude of one element's effect on another (Wilkinson & Friendly, 2009). The heatmap developed for this project was designed to serve as a heuristic tool and framework for helping DSCA prioritize technologies and task areas for further investigation and potential development of tailored business strategies.

Heatmap Methodology

To build the heatmap, the project team first created a technology taxonomy. The taxonomy draws from DoD documentation, industry reports, and academic publications to ensure inclusion of commercial technologies with military applications (i.e., dual-use). The taxonomy was reviewed by technical subject matter experts within IDA and outside research organizations. The project team further refined the technology taxonomy by excluding technologies that failed to meet the following criteria in the context of typical moderately capable partners:

1. **Applicability.** Technologies must be relevant to measurably improving and/or enabling a partner nations' defense and deterrence of larger aggressors.
2. **Sustainability.** There must be a robust or rapidly growing commercial market for the technology. For defense-unique technologies that have little to no commercial market, there must be a viable path to ensuring its continued production and sustainability. Examples include significant adoption and/or use by allied forces and/or plans by the DoD to acquire and scale the technology.
3. **Maturity.** Technologies must be technology readiness level (TRL) seven or higher.[7] This indicates the technologies are no longer experimental and have been successfully

[7] Note that TRLs are not typically used by commercial technology companies. The IDA team applied the TRL nomenclature because it is commonly used in and understood by the DoD and other government agencies. TRL 7 means that there has been a system prototype demonstration in an operational environment (Office of the Under Secretary of Defense for Research and Engineering, 2023).

prototyped, at a minimum, in a test reflective of expected operational conditions. The technology is generally ready for sale or transfer.

4. **Absorbability.** Those using the technology do not require specialized or advanced education (e.g., doctoral-level training). Training, certifications, and education on how to use the technology must be readily available (e.g., industry-provided) and generally consistent with what is commonly provided by the DoD to partner nations (e.g., via international military education and training [IMET]).[8]

Note, the taxonomy does not include technologies generally considered to be part of national infrastructure (e.g., telecommunications infrastructure such as fiber and 5G networks).

Next, the project team derived functional areas and associated tasks from the DoD's Universal Joint Task List (UJTL). UJTLs that may be applicable to moderately capable partners were included, whereas highly advanced or U.S.-only tasks were excluded. Examples of excluded functional areas and tasks are those pertaining to nuclear capabilities and top secret–level communications networks.

Finally, the project team coded the impact of every technology included on the taxonomy vis-à-vis functional areas and tasks. Impact was rated on a scale of high, moderate, low, or none:

Table 1. Impact Levels of Technologies in Taxonomy

Impact Level	Heatmap Color	Definition	Examples & Explanation
High Impact		The technology has a proven or expected ability to provide transformational capabilities and deliver outsized returns.	The technology can render an enemy capability obsolete and/or unlock or power the ability to use additional capabilities.
Moderate Impact		The technology has a proven or expected ability to improve or optimize how a task is executed.	The technology can improve important functions in task execution. However, the technology alone may not unlock the ability to leverage additional capabilities.
Low Impact		The technology has a proven or potential utility in supporting execution of a task.	The technology may be useful to how a task is executed, but may not be required or essential.
No Impact		The technology has no discernable application or direct relevance to a task.	N/A

Evaluating impact is inherently subjective. However, to mitigate bias in the impact analysis and ensure the validity of the impact scores, the IDA team conducted literature reviews, sought input from subject matter experts, and completed multiple internal reviews of the impact scores to ensure consistency and accuracy in scoring. Literature reviews included reviewing publications such as industry analyst reports (e.g., Gartner), scientific articles and journals, and open-sourced reporting on military use of commercial technologies. The IDA team collaborated with subject matter experts (e.g., technical experts, functional experts and leaders, and former industry executives) within and external to IDA to validate and adjust impact scores.

[8] IMET is a program that provides U.S. government funds for international military personnel to attend educational programs and training at U.S. military facilities.

Notwithstanding the use of multiple sources to determine impact, there will always be some degree of subjectivity in the heatmap. Further, drawing hard boundaries and distinctions between technologies is difficult as they often bleed together (e.g., AI is often a core component of modern software). For these reasons, the heatmap should not be interpreted as authoritative. However, it does provide a well-formed directional view of which technologies have the greatest potential to substantively improve or transform a partner nation's capabilities and thus informs DSCA's decisions on where to focus as it further develops strategies for delivering commercially available capabilities to partners.

Commercial Technology Trends

One approach to evaluating which technologies can yield the highest impact is to aggregate the impact scores of each technology (e.g., numerical scoring of 3, 2, 1, and 0 for high, moderate, low, and no impact). This approach points to five technology groupings with potential to deliver significant impact across a multitude of functional areas and tasks:

1. **Compute.** This includes cloud and edge computing. The heatmap indicates cloud computing has the greatest potential impact, including high impact on multiple command, control, communications, and computer systems (C4S) and intelligence tasks.
2. **Data.** All sub-categories of data (collection, processing, analytics, visualization, and management and storage) have the potential to deliver significant impact across a multitude of functional areas and tasks.
3. **Cyber.** Security of enterprise infrastructure, as well as applications, has relevance and impact on every task included in the heatmap.
4. **Multi-modal AI.** Multi-modal AI refers to AI and machine learning capabilities that can intake, process, and derive information from multiple types of data and from multiple sources (e.g., images and video ingested from different platforms). It can impact a variety of tasks, most notably intelligence, surveillance, and reconnaissance (ISR).
5. **Generative AI.** Generative AI has the potential to deliver impact across the vast majority of functional areas and tasks. Generative AI is a fast-developing technology and may have a larger impact on numerous tasks in the near future.

Beyond the top-five technologies, software-defined networking may also have a greater impact on some tasks, particularly C4S-related tasks. In addition, computer vision can have a large impact on intelligence-related tasks such as geospatial intelligence and ISR and a moderate impact on a multitude of tasks pertaining to force employment, sustainment, and C4S.

There are some technologies that may not impact or be useful to many task areas but can deliver outsized returns on specific capability areas and tasks. Not surprisingly, commercial satellite communications and commercial drones serve as useful examples. Commercial communications satellites have a proven ability to produce high impacts for C4S tasks and employment of joint fires and forces. Similarly, commercial UAVs are a high impact technology for targeting, ISR, and special operations (e.g., asymmetric warfare)—a finding amply demonstrated by commercial UAV use in Ukraine.

Impacts on Functional Areas and Tasks

In addition to understanding which technologies can produce the greatest impact, the heatmap offers a view into which functional areas and tasks could most benefit from using multiple commercial technologies. One task in particular has the greatest potential to benefit from integrating multiple commercial technologies together: ISR.

After ISR, the following tasks rate highest in terms of their potential to benefit from commercial technologies.

- Measurement and signals intelligence (MASINT)
- Maritime warning (i.e., maritime domain awareness)
- Joint command and control
- IT infrastructure
- Special operations[9]

The heatmap also illustrates that though some technologies have the potential to deliver higher impacts, realizing their transformational potential within a functional area or task often requires integrating multiple technologies and designing end-to-end solutions. For example, though commercial UAVs can highly impact ISR, the heatmap indicates that integrating commercial UAVs with cloud and edge computing; data collection, processing, and analytics; commercial communications and observation satellites; multi-modal AI and computer vision; and open-sourced software presents an opportunity to develop a truly transformational ISR capability for partner nations—evidenced by Ukraine's own experience integrating commercial drones with mobile apps, commercial satellite communications, and more to develop ISR, targeting, and joint fires capabilities.

Deep dives into each task area to understand how commercial technologies can be integrated into end-to-end solutions can offer partner nations and DSCA a clear roadmap for how to build transformative capabilities using commercial technology.

Next Steps and Recommendations

The following recommendations are organized across three phases DSCA can follow to best develop and execute new business strategies for delivering commercially available capabilities to moderately capable partners.

Phase 0: Pilot

As an initial phase (or in parallel with the other phases we outline in this section), DSCA should scope a pilot in partnership with a DIO, such as the DIU. A pilot will allow DSCA to test how engagement and procurement models typically employed by DIOs could be used to address partner nation needs. Scoping the pilot to a proactive use case rather than a response to acute needs in conflict will enable DSCA to pressure-test commercially available capabilities and learn how to best incorporate them into security assistance packages and long-term security cooperation strategies.

The pilot could be structured into two parts: the first being a solicitation (e.g., needs discovery and requirements generation, problem statement development, solicitation release) in partnership with one or more DIOs and the Department of Commerce, and the second (pending successful outcomes from the first) being identification of viable solutions for a partner nation.[10]

DSCA should consider selecting a technology or task area that is low-risk and for which there is deep institutional knowledge—both within the DoD and industry—for how to design and implement solutions. Of the low-risk task areas for which commercially available capabilities may be most impactful, distribution and logistics and health services may be the best candidates. DSCA might also focus attention on an underdeveloped, but important, capability area that is expected to be a future priority for moderately capable partners.

[9] Much like the technologies included in the taxonomy, some functional areas and tasks may overlap or be subsets of another. For example, MASINT can be considered a subset or specific type of ISR. IDA is continuing to refine the structure of the functional areas and tasks included in the heatmap to account for this.

[10] In some cases, DSCA and the larger security cooperation community may need to do some groundwork with certain partners so that they might understand the relevance of a class of technologies and its relevance to the partner security imperatives. Simply pointing out a set of solutions may not be effective for a partner who lacks an understanding of the potential opportunity space.

Phase 1: Market Validation and Market Fit

Engaging both industry and partner nations to understand their perspectives on strengths, weaknesses, opportunities, and threats associated with procuring commercially available capabilities will be critical to decision-making by DSCA's leadership, development of new strategies, and data-driven investments in new resources and capabilities within DSCA. For example, some industry partners may not be familiar with the arms export space and may need assistance to navigate U.S. and partner nation processes.

Recommendation 1.1. Define DSCA's value proposition in enabling sales of commercially available capabilities to partner nations.

DSCA needs to clearly define and articulate the value it brings to commercial transactions between industry and partner nations. As a starting point, IDA identified three potential benefits DSCA can offer to industry and partner nations:

1. **Customer Discovery and Requirements Generation.** DSCA can leverage deep relationships with partner nations and help partner nations develop commercial technology procurement strategies and roadmaps, informed by known and anticipated challenges and requirements.
2. **Convening Power and Network Effects.** DSCA's deep relationships with partner nations and key interagency partners (e.g., Department of Commerce and Department of State) may be compelling to industry partners interested in international sales and military applications of their technologies.
3. **Access to Funding**. It is possible that foreign military financing (FMF) could be used to facilitate direct commercial contracts (DCCs) of in-scope technologies. FMF may be useful for developing proofs of concept and pilots or for acquiring items for which there is little to no sustainment cost. DSCA could also explore other authorities that could be used to acquire commercial technologies.

Recommendation 1.2. Validate advantages and willingness among supply-side (industry) and demand-side (partner nation) stakeholders for DSCA to facilitate procurements of commercially available solutions.

DSCA should conduct private interviews and roundtable discussions to develop a holistic understanding of the concerns, objectives, and needs of potential supply-side stakeholders. DSCA should also leverage in-house market analysis and lessons from ongoing engagements and work with moderately capable partners to validate their interest and ability to procure, use, and sustain commercially available technologies.

Phase 2: Preparation and Launch

If Phase 1 indicates strategic feasibility and demand for DSCA to provide commercially available capabilities, DSCA should invest in the foundational resources and capabilities needed to develop and execute new business and go-to-market strategies. This may require additional appropriation and authority from Congress.

Recommendation 2.1. Establish a dedicated team focused on cultivating and facilitating need-driven engagements between industry and partner nations.

This team should be responsible for cultivating strategic partnerships inside the DoD and with industry and for working with SCOs and partner nations to determine needs and facilitate targeted engagements between partner nations and potential sources of commercially available capabilities.

Recommendation 2.2. Develop training materials, resources, and guidance for SCOs.

SCOs are not trained today on commercially available capabilities, nor does guidance exist for how to identify if and when a commercially available capability may be a viable solution—viability not being determined only by functionality, but also factors such as cost, complexity, and sustainability—for an identified partner nation requirement. The Defense Security Cooperation University should consider ways to prepare security cooperation officers to address commercially available capabilities when engaging with partners.

Recommendation 2.3. Enhance DSCA's market intelligence on capabilities and/or technologies of greatest potential impact.

Current market intelligence is critical to staying abreast of key innovations and market trends, understanding companies' offerings, and identifying new commercially available capabilities that may address partner nations' needs. DSCA could take the following steps to develop market intelligence:

Recommendation 2.3.1. Coordinate an industry outreach campaign to educate commercial technology providers on DSCA's mission, establish relationships, and build DSCA's brand awareness.

Send demand signals to industry by publishing DSCA's technology priorities, and create mechanisms through which industry can reach DSCA to share information on their products and offerings. The technologies priorities should be refined through information gathered during Phase 0 (pilot) and Phase 1.

Recommendation 2.3.2. Establish a consortium of industry partners who can advise and assist in the deployment of commercially available capabilities to partner nations.

DSCA can serve as a convener between industry and partner nations to facilitate customer discovery, systems architecture design, and procurements.

Recommendation 2.3.3. Subscribe to publicly available market intelligence platforms databases.

Acquiring access to commercial market intelligence platforms and databases can accelerate DSCA's visibility into technologies, companies, and industry trends.

Recommendation 2.3.4. Partner with DIOs and interagency offices with established market intelligence capabilities and procurement strategies focused on commercially available capabilities.

Establish relationships with DIOs and other interagency offices through which DSCA can submit targeted queries to identify commercially available capabilities that may meet partner nations' requirements. As noted earlier, DIU's interest in generating "impact through shared best practices, talent management, shared systems and processes, and enhanced teamwork" has potential to position DIU as an effective information aggregator and help streamline query requests for DSCA.

Phase 3: Execute and Scale

Recommendation 3.1. Conduct engagements with partner nations that can elucidate how they may apply commercial technologies in new, novel ways.

The heatmap reflects a U.S.-centric perspective of technology impact. Further, constant technological evolution and innovation may result in unforeseen applications of commercially available capabilities. Incorporating commercial technologies into exercises and wargames with partner nations will stimulate new ideas on ways to apply technologies and illuminate how partner nations can best harness them to execute critical tasks. With additional training and guidance, SCOs should also engage with partner nations on ways that commercially available technologies can meet their requirements and priorities.

Recommendation 3.2. Field assessment teams to apply the heatmap framework to specific partner nations and/or identify opportunities for commercially available capabilities to improve partner nations' abilities to execute critical tasks.

The heatmap presented in this report was developed without grounding or insights into how specific partner nations are currently using commercial technologies or executing critical tasks. Fielding assessment teams with combined expertise in commercial technologies, systems architectures, and priority task areas can produce data-driven insights into opportunities and risks associated with incorporating commercially available technologies into how partners execute priority tasks. DSCA in-house foreign market intelligence capabilities may provide initial insights on current and likely future partner demands.

Recommendation 3.3. Collaborate with partner nations and industry to develop tailored strategies and roadmaps for effective technology deployment, including building end-to-end solutions and transformative capabilities.

Technologies are often reliant on one another; helping partner nations incorporate commercially available capabilities into their operations requires a holistic understanding of interplay and interdependencies of technologies. Additionally, the means by which one commercially available capability is best delivered may not be the same means by which another is delivered. Strategies should reflect unique timelines, contractual mechanisms, upfront costs, and long-term costs associated with acquiring a given technology or solution. As noted above, some partners may need assistance in recognizing the opportunities through commercially available military capabilities and the demands of absorbing, applying, and sustaining procured solutions.

Additional Considerations

Other factors warrant examination, as they may impact the viability and efficacy of DSCA helping partner nations acquire and incorporate commercially available capabilities into their operations. Potential areas of continued research include:

- **Export controls.** Some technologies in the heatmap, including those that could significantly improve partner nations' capabilities, are subject to export controls.[11] It will be critical for DSCA to carefully study priority technologies to understand blockers, pathways, and requirements for helping partner nations acquire U.S. produced technologies that fall under export control restrictions. Interagency coordination and collaboration will likely be a core component of overcoming export-related blockers (and more).
- **Partner nations' culture of innovation and technical capacity.** There may be limitations to the lessons that can be drawn from Ukraine, particularly regarding the feasibility of helping other partner nations leverage commercially available capabilities for self-defense and deterrence. Ukraine's ability to innovate with commercial technologies may be due, in large part, to country's robust IT industry and sizeable technical workforce.[12] Further research should explore the importance of a culture of innovation for effective use of commercially available capabilities and ways to improve or encourage innovation within partner nations.

---

[11] See information on restrictions posted by the Bureau of Industry and Security (U.S. Department of Commerce, n.d.).
[12] By some estimates, Ukraine's IT workforce totals 285,000 specialists who generate 4% of the national GDP (~$6.8 billion). Of that population, many are in the armed forces or work in national cyber defense (Kontsevoi, 2022; Tan, 2022).

- **DSCA and SCO culture of innovation**. The DoD is cultivating new approaches, mindsets, and human capital strategies for considering problems, engaging industry, innovating rapidly, attracting technical talent, and more. To help partner nations leverage commercially available capabilities, the cultural shifts the DoD is cultivating in the defense acquisition community will need to extend to DSCA and SCOs.
- **Risk mitigation**. Prior to encouraging or facilitating a sale, DSCA should have a well-informed understanding of how commercially available capabilities could be used by partner nations and the potential risks of such use.
- **Procurement methods**. As noted above, FMF may present an option for facilitating procurements, but its utility is limited. Procurement methods must be timely as commercial technology evolves rapidly. Further research should include a careful examination of existing authorities, consideration of potential needs for new authorities, and engagement with both partner nations and industry to understand their procurement-related needs.

Conclusion

Commercially available technologies present significant opportunities for helping partner nations improve capabilities and also significant challenges. The IDA team continues to research this topic, and we appreciate the opportunity to present our findings to date at the 21st Annual Acquisition Research Symposium. We look forward to your feedback and engagement.

References

Additions of Entities to the Entity List and Removal of Entity from the Entity List, 15 C.F.R. § 744 (2023). https://www.federalregister.gov/documents/2023/06/14/2023-12726/additions-of-entities-to-the-entity-list-and-removal-of-entity-from-the-entity-list

Additions and Revisions to the Entity List and Conforming Removal from the Unverified List, 88 Fed. Reg. 38739 (December 19, 2022). https://public-inspection.federalregister.gov/2022-27151.pdf

Anwar, N. (2023, February 7). *World's largest drone maker is unfazed—even if it's blacklisted by the U.S.* CNBC. https://www.cnbc.com/2023/02/08/worlds-largest-drone-maker-dji-is-unfazed-by-challenges-like-us-blacklist.html

Austin, L. (2022, October 27). *2022 national defense strategy*. U.S. Department of Defense. https://media.defense.gov/2022/Oct/27/2003103845/-1/-1/1/2022-NATIONAL-DEFENSE-STRATEGY-NPR-MDR.PDF

AUVSI. (2022). *Pricing—Uncrewed systems and robotics database.* https://roboticsdatabase.auvsi.org/pricing

AUVSI. (n.d.). *Uncrewed systems & robotics database*. https://www.auvsi.org/usrd

Beck, D. A. (2024). *"DIU 3.0" Scaling Defense Innovation for Strategic Impact*. Center for a New American Security. https://www.cnas.org/publications/reports/diu-3-0

Borger, J. (2022, December 18) "Our weapons are computers": Ukrainian coders aim to gain battlefield edge. *Guardian.* https://www.theguardian.com/world/2022/dec/18/our-weapons-are-computers-ukrainian-coders-aim-to-gain-battlefield-edge

Bouey, J., Hu, L., Scholl, K., Marcellino, W., Dossani, R., Malik, A., Solomon, K., Zhang, S., & Shufer, A. (2023). *China's AI exports: Technology distribution and data safety*. RAND. https://www.rand.org/pubs/research_reports/RRA2696-2.html

Cheng, E. (2023, March 5). *China to increase defense spending by 7.2%.* CNBC. https://www.cnbc.com/2023/03/05/china-defense-budget-two-sessions.html

China Power Team. (2018). *How dominant is China in the global arms trade?* CSIS. https://chinapower.csis.org/china-global-arms-trade/

Defense Innovation Unit. (2023). *DIU's FY22 year-in-review*. https://www.diu.mil/fy22-year-in-review

DoD Innovation Pathways. (n.d.). *Innovation organizations*. https://www.ctoinnovation.mil/innovation-organizations/

Google. (n.d.). *Ukraine war drone incidents 2024*. Google Sheets. https://docs.google.com/spreadsheets/d/1oltrQ7RceC8w1eR2ttpoqSz3zHhW1-tZkrU7yfZTqAU/edit#gid=0

Gosselin-Malo, E. (2023). *Ukraine continues to snap up Chinese DJI drones for its defense*. C4ISRNet. https://www.c4isrnet.com/global/europe/2023/10/23/ukraine-continues-to-snap-up-chinese-dji-drones-for-its-defense/

Greenwood, F. (2023). The drone war in Ukraine is cheap, deadly, and made in China. *Foreign Policy*. https://foreignpolicy.com/2023/02/16/ukraine-russia-war-drone-warfare-china/

Jones, G., Egan, J., & Rosenbach, E. (2023). *Advancing in adversity: Ukraine's battlefield technologies and lessons for the U.S.* Harvard Belfer Center. https://www.belfercenter.org/publication/advancing-adversity-ukraines-battlefield-technologies-and-lessons-us

Konkel, F. (2022). *Ukraine tech chief: Cloud migration "saved Ukraine government and economy."* Nextgov. https://www.c4isrnet.com/global/europe/2023/10/23/ukraine-continues-to-snap-up-chinese-dji-drones-for-its-defense/

Kontsevoi, B. (2022, October 12). *The Ukrainian IT industry is alive and healthy*. Forbes. https://www.forbes.com/sites/forbestechcouncil/2022/10/12/the-ukrainian-it-industry-is-alive-and-healthy/?sh=5dd8f2d67f2c

Kynge, J., Hopkins, V., Warrell, H., & Hille., K. (2021). Exporting Chinese surveillance: The security risks of "smart cities." *Financial Times.* https://www.ft.com/content/76fdac7c-7076-47a4-bcb0-7e75af0aadab

McDonald, J. (2023). *China restricts civilian drone exports, citing Ukraine and concern about military use.* AP News. https://apnews.com/article/china-ukraine-russia-drone-export-dji-e6694b3209b4d8a93fd76cf29bd8a056#:~:text=BEIJING%20%28AP%29%20%E2%80%94%20China%20imposed%20restrictions%20Monday%20on,says%20it%20is%20neutral%20in%20the%2017-month-old%20war

Montgomery, M., & Sayers, E. (2023). Don't let China take over the Cloud—US national security depends on it. *The Hill.* https://thehill.com/opinion/national-security/4307002-dont-let-china-take-over-the-cloud-us-national-security-depends-on-it/

Morris, F. (2023, April 14). *Slow manufacturing and price gouging threaten the new U.S. military arms race*. NPR. https://www.npr.org/2023/04/07/1168725028/manufacturing-price-gauging-new-u-s-military-arms

Nagar, S. (2022). ZTE's revenge: Russia's technological power vacuum in the wake of the Ukraine war. *Harvard International Review.* https://hir.harvard.edu/ztes-revenge-russias-technological-power-vacuum-in-the-wake-of-the-ukraine-war/

National Urban Security Technology Laboratory. (2022, January). *Saver technote: Laser protective eyewear*. U.S. Department of Homeland Security. https://www.dhs.gov/sites/default/files/2022-01/SAVER_Laser%20Protective%20Eyewear_TechNote_508%20final_18Jan2022.pdf

Nouwens, M., & Legarda, H. (2018). *China's pursuit of advanced dual-use technologies.* International Institute for Strategic Studies. https://www.iiss.org/research-paper/2018/12/emerging-technology-dominance

Office of the Under Secretary of Defense for Research and Engineering. (2023). *Technology readiness assessment guidebook*. https://www.cto.mil/wp-content/uploads/2023/07/TRA-Guide-Jun2023.pdf

Russel, D., & Berger, B. (2020). *Weaponizing the belt and road initiative.* Asia Society Policy Institute. https://asiasociety.org/sites/default/files/2020-09/Weaponizing%20the%20Belt%20and%20Road%20Initiative_0.pdf

Sahin, K. (2020). *The West, China, and AI surveillance.* Atlantic Council. https://www.atlanticcouncil.org/blogs/geotech-cues/the-west-china-and-ai-surveillance/

Syamsudin, A. (2023). *Huawei's role in Indonesia raises digital colonization concerns.* Radio Free Asia. https://www.rfa.org/english/news/china/china-bri-indonesia-09272023104442.html

Tan, H. (2022, April 8). *Ukraine's 285,000 IT specialists power apps and software around the globe, and many of them are still working from Ukraine as the war rages around them*. Business Insider. https://www.businessinsider.com/ukraine-it-specialists-still-working-through-war-2022-4

U.S. Department of Commerce. (n.d.). *Export administration regulations*. Bureau of Industry and Security. https://www.bis.doc.gov/index.php/regulations/export-administration-regulations-ear

U.S. Department of Commerce Industry & Analysis-Aerospace Office and U.S. Commercial Service. (2024, February 26). *Aerospace and defense exporter alert, February 2024*. International Trade Administration. https://content.govdelivery.com/accounts/USITATRADE/bulletins/38a8c39

U.S. Department of Defense. (2023). *Military and security developments involving the People's Republic of China*. https://media.defense.gov/2023/Oct/19/2003323409/-1/-1/1/2023-MILITARY-AND-SECURITY-DEVELOPMENTS-INVOLVING-THE-PEOPLES-REPUBLIC-OF-CHINA.PDF

U.S. Department of Homeland Security. (n.d.-a). *Blue UAS for first responders*. https://www.dhs.gov/science-and-technology/saver/blue-uas-first-responders

U.S. Department of Homeland Security. (n.d.-b). *National urban security technology laboratory*. https://www.dhs.gov/science-and-technology/national-urban-security-technology-laboratory

U.S. Department of Homeland Security. (n.d.-c). *System assessment and validation for emergency responders (SAVER) program*. https://www.dhs.gov/science-and-technology/saver

Weinbaum, C., O'Connell, C., Popper, S., Bond, S., Byrne, H., Curriden, C., Weider Fauerbach, G., Lilly, S., Mondschein, J., & Schmid, J. (2022). *China's defense industrial base.* RAND. https://www.rand.org/pubs/research_briefs/RBA930-1.html

Wilkinson, L., & Friendly, M. (2009). The history of the cluster heat map. *The American Statistician*, *63*(2), 179–184. https://doi.org/10.1198/tas.2009.0033

PANEL 16. INNOVATIVE CONTRACTING APPROACHES IN DOD

Thursday, May 9, 2024	
10:30 a.m. – 11:45 a.m.	**Chair: Major General Alice W. Trevino, USAF,** Deputy Assistant Secretary for Contracting, Office of the Assistant Secretary of the Air Force for Acquisition, Technology and Logistics ***Partnership Intermediary Agreements: Analysis of Effective Practices in the Air Force Global Strike Command*** Rene Rendon, Naval Postgraduate School ***Commercial Solutions Openings – Innovative and Impactful*** E. Cory Yoder, Naval Postgraduate School ***The Efficacy of Optimized Government-Industry-Academia Co-Education for Major Weapon Systems Cost/Price Analysis and Contract Negotiations*** Kelley Poree, Naval Postgraduate School

Major General Alice W. Trevino, USAF—serves as the Deputy Assistant Secretary for Contracting, Office of the Assistant Secretary of the Air Force for Acquisition, Technology and Logistics, the Pentagon, Arlington, Virginia. She is responsible for all aspects of contracting relating to the acquisition of weapons systems, logistics, operational and enterprise efforts for the Air Force and provides contingency contracting support to the geographic combatant commanders. She leads a highly skilled staff of mission-focused business leaders and acquisition change agents to deliver $825 billion in United States Air Force and Space Force platforms. Additionally, she is the Contracting Functional Manager for over 8,000 professionals, who execute programs worth $70 billion annually for the Department of the Air Force.

Maj. Gen. Treviño received her commission from the U.S. Air Force Academy in 1993 and is a joint qualified officer with extensive deployment experience in support of combat, humanitarian and peacekeeping/enforcement operations to Croatia, Turkey, Oman, Kuwait and Afghanistan.

Prior to her current assignment, she was the Commander of the Air Force Installation Contracting Center. Maj. Gen. Treviño has also served as the Deputy Secretary of Defense’s Principal Military Assistant,; unlimited dollar warranted procuring Contracting Officer for major defense programs; and the Senior Contracting Official-Afghanistan for U.S. Central Command. She has commanded two Air Force units at the squadron level, joint units at the group and wing levels, and an Air Force unit at the wing level.

Partnership Intermediary Agreements: Analysis of Effective Practices in the Air Force Global Strike Command

Allan E. Cameron—is a distinguished and results-driven acquisition professional with over 11 years of expertise in defense procurement, strategic leadership, and policy development. Cameron graduated from the U.S. Air Force Academy with a Bachelor of Science in Economics and earned his Masters of Business Administration from the Naval Postgraduate School. Throughout his career, Cameron has led at the operational, enterprise, and systems contracting levels in the USAF. Most recently, Cameron served as the Deputy Division Chief of Policy at the Global Strike Command Contracting Operating Location for a portfolio valued over $323 million. Cameron also served as the Lead Agreements Officer and administered an $85 million Partnership Intermediary Agreement within Global Strike Command for technological development and innovation. [allan.cameron.v@gmail.com]

Dr. Rene G. Rendon—is an Associate Professor at the Naval Postgraduate School, where he teaches graduate defense acquisition and contract management courses. He completed a career as an Acquisition Contracting Officer in the USAF, serving as a Contracting Officer for weapon systems such as the Peacekeeper ICBM, Maverick Missile, and the F-22 Raptor. He also served as a Contracting Squadron Commander and Director of Contracting for the Space Based Infrared Satellite program and the Evolved Expendable Launch Vehicle program. Rendon has published in the Journal of Public Procurement, the Journal of Contract Management, and the Journal of Purchasing and Supply Management. [rgrendon@nps.edu]

Abstract

One challenge for the Department of Defense (DoD) in using innovative approaches to acquisition includes identifying best practices, lessons learned, and successful use cases and disseminating this information to other agencies for adoption. The literature includes successful use cases for Other Transaction Authorities (OTAs), but the use of Partnership Intermediary Agreements (PIAs) is still comparatively limited. The purpose of this research is to discuss effective practices in the use of PIAs within the U.S. Air Force. Specifically, our research focuses on the Air Force's Global Strike Command PIA with the Cyber Innovation Center (CIC) for support and research within the nuclear deterrence and operations domain. In this research, we answer the question, "What are the most effective practices in the use of PIAs that will foster innovation, maximize collaborative outcomes, and achieve considerable return on investment?" We draw from practical applications in the CIC as well as other reports from the literature. Although this research focus is limited to the Air Force's Global Strike Command PIA with the CIC, our findings can still be applied to similar PIAs in the DoD that are focused on acquiring innovative, agile, and more collaborative solutions in a timelier manner.

Introduction

The U.S. Department of Defense (DoD) obligates billions of dollars annually for the procurement of goods and services that are essential for homeland defense. The majority of these procurements are structured using contracts based on the Federal Acquisition Regulation (FAR) and its related supplements; however, some agencies are authorized to use non-FAR instruments as a way of acquiring innovative, agile, and more collaborative solutions in a timelier manner. The use of Partnership Intermediary Agreements (PIAs) is still in its early adoption phase, yet the DoD has already experienced successful acquisition outcomes using these non-FAR instruments. Challenges for the DoD in using these innovative approaches to acquisition include identifying best practices, lessons learned, and successful use cases for its non-FAR contracts and disseminating this information to other agencies for adoption.

The purpose of this research is to discuss effective practices in the use of PIAs within the U.S. Air Force. Specifically, this research focuses on the Air Force's Global Strike Command PIA with the Cyber Innovation Center (CIC) for support and research within the nuclear deterrence and operations domain. In this research, we answer the question, “What are the most effective practices in the use of PIAs that will foster innovation, maximize collaborative outcomes, and achieve considerable return on investment?” We draw from practical applications in the CIC as well as other reports from relevant acquisition literature. Our research and identified findings will benefit other Air Force PIA initiatives as well as PIAs planned and implemented throughout the DoD.

Background

PIAs are nontraditional procurement instruments established across the federal government. A PIA is authorized and governed by congressional statute that establishes its legal framework. The congressional statutes serve as the legal framework and basis for the government to authorize use of a PIA in public procurement. In establishing a PIA, the primary intent for the government is to drive innovation and foster collaboration with the Partnership Intermediary (PI) through technology transfer, licensing, and data rights (Defense Acquisition University, 2020).

The congressional statutes for PIAs are as follows: 15 U.S.C. § 3715, 15 U.S.C. § 3705, 15 U.S.C. § 3707, and 10 U.S.C. § 4124 (Defense Acquisition University, 2020). Specific to the armed forces and renumbered from 10 U.S.C. § 2368, 10 U.S.C. § 4124 tasks the secretary of each military department to “reengineer management and business processes and adopt best-business and personnel practices at the Centers of the Secretary concerned in connection with the capability requirements of the Centers, so as to serve as recognized leaders in such capabilities throughout the Department of Defense and in the national technology and industrial base” (Centers for Science, Technology, and Engineering Partnership, 2023).

The statute authorizes the use of non–FAR based agreements for defense procurement contracting officers to award within the context of best practices through their respective centers (e.g., designated federal laboratory). Additionally, 10 U.S.C. § 4124 Subsection (f) outlines the “use of partnership intermediaries to promote defense research and education” through a contract, “memorandum of understanding or other transaction with a partnership intermediary to perform services for the Department of Defense that increase the likelihood of success in the conduct of cooperative or joint activities of the Center with industry or academic institutions” (Centers for Science, Technology, and Engineering Partnership, 2023).

The principal intent of the federal statutes within defense procurement is to promote partnerships with academia, nonprofit organizations, and small businesses in order to drive an outcome of enhanced research and development, innovation, technology transfer, and collaboration for a more sustainable and cost-effective defense acquisition ecosystem.

Methodology

This report provides an explanatory and qualitative analysis of publicly available sources, including but not limited to PIAs, congressional statutes, the Institute for Defense Analyses, the Government Accountability Office, the RAND Corporation, the Congressional Research Service and congressional hearings, as well as the current PIA landscape within the defense sector. The analysis blends existing literature, official documents and reports, and studies to draw inferences and conclusions on the most effective uses of the PIA. Additionally, as one of the authors of this study has 11 years of recent and relevant practitioner experience, this study is able to synthesize direct insights acquired from fieldwork with academic analysis.

The research methodology includes a comprehensive examination of current congressional statutes, government reports, legislature subcommittee hearings, and third-party research institutions. Furthermore, the methodology employs a case study–based approach to the specific success within Air Force Global Strike Command. The analytical techniques employed in the methodology are as follows: policy analysis, content review, and thematic analysis through categorical grouping of relevant and publicly available literature.

Findings and Analysis

The qualitative analysis performs a research-oriented approach on the existing literature, studies, and case-based Global Strike Command PIA to present findings on the most effective practices for the use of PIAs. The intent of the research is to capitalize on nontraditional acquisition practices from a PIA that foster innovation, maximize collaborative outcomes, and achieve considerable return on investment within the Department of the Air Force. The primary findings from the qualitative analysis identified the following most effective practices for the use of PIAs: effective sharing of information, standardized administrative contracting procedures, proximity of location for partnership intermediary, aligned performance evaluation metrics, and lessons learned and after-action reports. These are further discussed in the next section.

Effective Sharing of Information

Effective communication and collaboration are essential for the successful outcome and effectiveness of PIA execution. According to an Institute for Defense Analyses report, effective communication and collaboration between not only the government and the PI but also other PIs increases the likelihood of usable technologies and more intertwined collaboration between multilateral organizations (Peña et al., 2020).

Within Global Strike Command, the PI provides a platform through their collaborative environment located within the Cyber National Park in Shreveport, LA, where the government effectively communicates with both the PI and nontraditional vendors. In these instances, the collaborative environment provides a “brick and mortar” location for both contractors and the government to actively collaborate on ongoing innovative projects under the PIA. In addition to utilization of a collaborative environment, the government has infused the PI into the Science & Technology Department on Barksdale Air Force Base (AFB). By leveraging these capabilities, the PI and the government enhance their overall ability to transfer information as well as effectively communicate through direct contact with continual feedback in real time (Department of the Air Force, 2023).

The key recommendation from the research indicates that prototyping and fielding of technology derived from a PIA must provide continuity between stakeholders through turnover at each stage of the acquisition process to ensure transfer of innovative technological outcomes as well as government usability on future acquisitions resultant in effective cost-savings and return on investment. Through study, practice, and theory, the importance of effective communication between members of both the government and the PI acquisition team is paramount for successful implementation and administration of requirement activities in a PIA.

Standardized Administrative Contracting Procedures

PIAs have often been traditionally underutilized within the DoD due to the lack of both readily available information as well as knowledge on many non-FAR acquisition processes (Dunn, 2022). Within the past decade, the DoD’s increased interest in nontraditional procurement instruments to achieve rapid innovation with effective collaboration has driven the need for simplified and standardized procedures (Vergun, 2021).

From a practitioner’s perspective at Global Strike Command, administrative and contracting procedures can lead to some of the most time-consuming delays for operational execution under a PIA. For example, Collaborative Project Designs (CPDs) under the

Collaborative Project Order (CPO) often have significantly varying specifications relative to their scope and intended outcome. As opposed to similar specifications, typical for Task and Delivery Orders under IDIQ contracts under FAR procedures, the wide breadth of differentiation for CPDs under a CPO can create unfamiliarity when administering contracting procedures. Ultimately, the unfamiliarity can lead to “noise interference” decision-making, miscommunication with acquisition team members, and even delayed award by seasoned government acquisition professionals (Peña et al., 2020).

While the recommendation for a standardized process is not to apply a “one-size fits all” approach, the rationale for streamlining and simplification is to provide a transparent roadmap through the duration of the PIA’s life cycle across the contract life-cycle phases, such as pre-award, award, and post-award, and technology transfer outcomes. In understanding the challenges of DoD policy and human resource activity, the government could achieve a more efficient and effective approach when administering a PIA. By incorporating these recommendations, this article finds that the government would have a higher likelihood to achieve efficient administrative implementation and execution as well as significantly increased cost savings. Furthermore, the recommendation would provide for improved strategic use of resource allocation with an expectation for deployable technological outcomes within the agency (Dunn, 2022).

Proximity of Location for Partnership Intermediary

This research finds that the proximity of a location can be a principal measure in gauging the effectiveness of a PIA. A study performed by the Institute for Defense Analyses surveyed 106 procurement professionals, 48 from 33 DoD entities and 58 from 28 different PIs, who responded to semi-structured interviews as well as provided input based on their expertise of PIAs (Peña et al., 2020). A primary consideration of the study was the efficient and effective use of PIs co-located with their administering DoD entity. In multiple situations, the study concluded that proximity led to enhanced technological outcomes within the collaborative environment by co-locating both DoD employees and their respective PI contractor counterparts.

From a practitioner’s perspective, close proximity has been instrumental in facilitating technology transfers and enhancing collaborative environments between the PI and the government. As examined at the Global Strike Command PIA in Shreveport, LA, the PI is located 17 miles from Barksdale AFB. The CIC serves as the Global Strike Command PI. The CIC is a hub for research and development, prototyping, as well as hosting various Industry Days (Academic Partnership Engagement Experience, 2021). Since the inception of the Global Strike Command PIA, both the PI and the government have reported a significant return on investment to the Air Force. These figures include an annual return of $4 million as well as $200 million in indirect cost savings directly attributable to a plethora of sourced technologies derived from the collaborative environment in Shreveport, LA (Department of the Air Force, 2023).

Our research recommends that proximity is one of the driving factors that lead to facilitation of collaborative and more efficient outcomes regarding problem identification, technological methodology and solutions, and overall cost savings (Peña et al., 2020).

Aligned Performance Evaluation Metrics

Our findings indicate that continuous measurement, tracking, and analysis of the performance of a PIA is critical for optimization of efficiency and effectiveness. By tracking the progress and performance of individual projects (e.g., CPDs), the government and the PI can develop a structured collaborative environment. Key stakeholders for meetings at regular intervals consist of acquisitions team program managers, finance officers, and contracting officers as well as their PI counterparts, which enhances the timelines, effectiveness, progress, and overall performance of the PIA (Peña et al., 2020).

Performance evaluation and compliance are key components for measuring the success of a PIA. While many project designs of a PIA might not lead to a readily deployable end product, the technology derived through the duration of the CPD should be monitored for defense uses, modifications, and progress of development (Peña et al., 2020). CPDs experience a high propensity of variability and degree of change. As many project designs do not directly correlate to a specific solution, project designs can lead to alternative and nontraditional solutions for an executed problem statement. In the Global Strike Command PIA, the government and the PI have maintained continual awareness within the program and acquisition team, which led to effectively maintaining schedule, milestones, and performance accountability on individual projects with high returns on investment.

Our research recommends that the government acquisitions teams (especially the agreements/contracting officer) must maintain awareness and knowledge of tracking and reporting metrics of specific CPDs in conjunction with the PI. Furthermore, the team should monitor and collaborate on the PI's monthly and quarterly financial compliance statements, performance-oriented outcome meetings, and end-state technology transfer summarization meetings. By understanding the operational necessity, performance evaluation becomes more manageable and less labor-intensive.

Lessons Learned and After-Action Reports

Lastly, our research finds that both lessons learned and after-action reports are essential post-award requirements necessary to gauge the effectiveness and outcome of the PIA as well as refine future project design implementation and practices for use. By detailing thorough after-action reports and lessons learned on outcomes of associated project designs, the government team can make more informed decisions when constructing future CPD problem statements.

Non-traditional contractors have provided the Global Strike Command PIA with multiple, varied solutions to problem statements, and the PIA has significantly benefited from employing these after-action reports and lessons learned as advanced market and concept research for future acquisitions. One of the first project design sprints hosted by the Global Strike Command PIA was the Emergency Aircrew Response (EAR) Readiness, which partnered with the Air Force Nuclear Weapons Center (AFNWC) and AFWERX. The PI under the Global Strike Command PIA, CIC, led a Challenge Day that selected three finalists out of 47 total submissions for an EARs solution on legacy mass aircrew alerting systems across Barksdale AFB. Of the three finalists selected to participate in prototyping the developed technologies, the PI provided detailed analysis on the lessons learned, which led to more efficient practices being implemented on future CPDs at Global Strike Command. The outcome of the project design sprint contributed to the indirect cost savings to the government of $200 million (Department of the Air Force, 2023). Furthermore, the outcome led to direct, non-traditional contractor engagement with primary defense contractors on large-scale weapon system acquisitions, specifically at Hanscom AFB.

It is recommended that the government and the PI establish after-action reports, metrics, and deliverables to evaluate the success of PIAs. The Air Force and other military departments would benefit from an after-action report that includes milestones, outcomes, data/licensing rights, and deliverables. This report would provide a tangible indication of the estimated return on investment less the cost of the PI-sourced contractor as well as the award amount to the sourced contractor for any follow-on programs. Lastly, the report could provide enhanced awareness of technologies not readily available within the defense industry.

Conclusion

The transformative environment of the modern defense procurement landscape leads to a pressing need for the ability to harness innovation in order to safeguard national security interests while maintaining technological relevance within the Department of the Air Force. Within this context, the Department of the Air Force is increasingly recognizing the importance of nontraditional tools such as PIAs within its collaborative acquisition framework. As the research highlights, these nontraditional procurement instruments have proven both effective and efficient in creating links between the DoD as well as the advancement of technology from the private sector.

Our research identified several tenets that are integral to optimizing the use of PIAs. By emphasizing effective information-sharing, the Air Force Global Strike Command is able to use that transparency to ensure all stakeholders are synchronized in both their objectives and efforts outlaid in the requirements of the PIA. Standardized administrative contracting procedures derive the need for consistency, efficiency, and predictability in PIA execution. Standardizing the procedures not only streamlines the process but also ensures trust in and understanding of both the government and the PI. Additionally, the study emphasized the key importance of maintaining strategic proximity between both PIs and their supported government counterparts. By ensuring a geographically accessible PI, the real-time collaboration led to a more organic and immediate exchange of ideas and solutions for various projects under the PIA. Furthermore, the performance evaluation metrics finding displayed the efficiency, success rate, and outcome of a collaborative and results-oriented approach in line with the key defense objectives of the overall acquisition project. Potentially one of the most engaging findings is the emphasis on Lessons Learned and After-Action Reports, which might seem readily apparent. However, underscoring the importance of these findings within the Department of the Air Force, lessons learned and after-action reports provide a tool for continual feedback improvement as well as analyzing past actions to refine future strategies for the PIA or follow-on acquisition. By employing these iterative and reflective tools, the government team can either replicate or eliminate processes from specific projects within the PIA.

Our research's primary limitation of the findings is that the Global Strike Command PIA was the primary case study utilized for analysis from a practitioner perspective. While analyzed in conjunction with existing literature, the categorical findings were restricted to the primary case study. The research's contribution is to highlight existing best practices from a practitioner's perspective intertwined with qualitatively relevant research literature. Future research could incorporate interviews and perspectives from agreements officers, program managers, and PIs on alternative PIAs in different MAJCOMs across the Air Force.

References

Academic Partnership Engagement Experience. (2021, July). *What are Partnership Intermediary Agreements (PIAs) and how can they help small businesses and academia innovate and commercialize technology*. https://apex-innovates.org/news/blog/what-are-partnership-intermediaries-agreements-pias-and-how-can-they-help-small

Air Force Global Strike Command. (2020, May). *Collaborative environment drives rapid solutions for global strike, warfighters*. https://www.afgsc.af.mil/News/Article-Display/Article/2191971/collaborative-environment-drives-rapid-solutions-for-global-strike-warfighters/

Centers for Science, Technology, and Engineering Partnership, 10 U.S.C. § 4124 (2023). https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title10-section4124&num=0&edition=prelim

Defense Acquisition University. (2020). *Contracting cone: Partnership Intermediary Agreement (15 U.S.C. § 3715).* https://aaf.dau.edu/aaf/contracting-cone/rd-agreements/pia/

Department of the Air Force. (n.d.). *Department of the Air Force technology transfer and transition: Partnership Intermediary Agreements (PIAs)*. Retrieved July 4, 2023, from https://www.aft3.af.mil/PIAs/PIAs-List/

Department of the Air Force. (2023). *Department of the Air Force technology transfer and transition: Annual report 2022* (AFRL-2023-0887). https://media.defense.gov/2023/Feb/24/2003167320/-1/-1/0/DAFT3-FY22_ANNUALREPORT-FINAL_AFRL-2023-0887%20(1).PDF

Dunn, R. (2022). *The other transactions & partnership intermediary connection*. Strategic Institute. https://strategicinstitute.org/other-transactions/ot-pia-connection/

GAO. (2017, July). *Military acquisitions: DoD is taking steps to address challenges faced by certain companies* (GAO-17-644). https://www.gao.gov/assets/690/686525.pdf

GAO. (2020, October). *Army modernization: Army should improve use of alternative agreements and approaches by enhancing oversight and communication of lessons learned* (GAO-21-8). https://www.gao.gov/assets/gao-21-8.pdf

Peña, V., Mandelbaum, J., Babou, T. F., & Cavanaugh, C. P. (2020). *Opportunities to advance Department of Defense technology transfer with Partnership Intermediary Agreements*. Institute for Defense Analyses. http://www.jstoFr.org/stable/resrep36528

Vergun, D. (2021, October). *Investments in technology crucial to staying ahead, innovation unit director says*. DoD. https://www.defense.gov/News/News-Stories/Article/Article/2818899/investments-in-technology-crucial-to-staying-ahead-innovation-unit-director-says/

Commercial Solutions Openings – Innovative and Impactful

Mary Beth Colavito—is the Services Division Director and a Contracting/Agreements Officer with an unlimited warrant in the Contracts Management Office of the Defense Advanced Research Projects Agency (DARPA). She oversees a $2.5 billion ceiling portfolio with more than 100 active contracts/orders. She graduated with her Bachelor of Science from the University of Maryland–College Park and her Master of Science in Contract Management from the Naval Postgraduate School.

Eric W. Washburn— is a graduate of the Naval Postgraduate School and a Branch Chief with Air Force Test Center at Eglin Air Force Base, Florida. Washburn serves as Unlimited Contracting Officer, Agreements Officer, and Branch Chief for the Enterprise Services and Innovative Solutions Contracting Branch, managing a diverse portfolio of contracts critical to the national defense including Electronic Warfare Test Systems, Hypersonics, Artificial Intelligence, and F-35 support.

E. Cory Yoder—is a senior lecturer/professor at the Naval Postgraduate School. He holds an M.A. in National Security and Strategic Studies from the Naval War College, an M.S. in Management with a concentration in contract management, a B.S. in Business Management from Indiana University, and a Business Resource Management certificate from the University of Virginia–Darden. Yoder is a Beta Gamma Sigma honor society member and Rear Admiral John J. Schieffelin awardee.

Abstract

This research provides an analysis of the Department of Defense (DoD)'s use of the Commercial Solutions Opening (CSO), a new general solicitation technique to acquire innovative solutions. The purpose of this research is to identify strengths, weaknesses, and best practices of CSOs and make recommendations based on those observations. This research also analyzes the statistical difference in the procurement lead times of contracts awarded from CSOs compared to Federal Acquisition Regulation (FAR)–based solicitations by conducting a statistical analysis of Federal Procurement Data System (FPDS) data. We reviewed data from CSO Cross Talks, congressional briefings and reports, and protest filings to identify 27 strengths, seven weaknesses, and 43 best practices for CSOs. These findings were then categorized by topic areas for systematic analysis. We developed eight recommendations focused on training and development, policy changes, and tracking and reporting, each with their anticipated benefits and methods to implement. As a solicitation technique, the CSO is a valuable tool to achieve innovation, but prudent planning and application of this research's identified best practices are critical to ensure acquisition success. By implementing the recommendations provided in this research, the DoD will be postured to utilize the CSO solicitation technique to its fullest potential, closing the technological capability gap and providing for better defense capabilities to the nation.

Introduction

It is no secret that the Department of Defense (DoD) traditional acquisition process is slow. For the purposes of this research, "traditional" is defined as Federal Acquisition Regulation (FAR)–based solicitation and award techniques. Since the 1990s, the acquisition process has appeared in some form in the list of top DoD challenges reported by the DoD Inspector General (IG) and has been called "inflexible" (Section 809 Panel, 2018, p. 6), "inefficient" (DoD Inspector General [DoDIG], 2015, p. 10), and "slow" (DoDIG, 2022, p. 7). In 2019, the U.S. Government Accountability Office (GAO) issued the report *DoD Acquisition Reform: Leadership Attention Needed to Effectively Implement Changes to Acquisition Oversight*, which discusses congressional concerns over DoD's weapons acquisition process, citing the processes' bureaucracy and delays in fielding innovations (GAO, 2019). This same report discusses the DoD's intent to increase the speed of the acquisition process through pursuing legislative reforms and acknowledges that the DoD has begun to execute those reforms, including realigning certain decision and oversight from the Office of the Secretary of Defense (OSD) to the subordinate military departments, as well as using more streamlined processes.

Regardless of these changes, the DoD still struggles to achieve rapid acquisition objectives, narrowing the strategic and defense capabilities gap between the United States and near-peer adversaries. Recent notable examples of this acquisition reform include the Middle Tier of Acquisition (MTA) Pathway for Rapid Prototyping and Rapid Fielding authorized by Section 804 of the Fiscal Year (FY) 2016 National Defense Authorization Act (NDAA, 2015), Awareness of Other Transaction (OT) Authority, and the adoption of industry standards in acquisition. Even with these reforms, the DoD acquisition process remains slow, expensive, and bureaucratic. In 2021, and in furtherance of rapid acquisition objectives, Congress codified Public Law 117-81, the NDAA for FY2022. Section 803 of the act provides permanent authority for a new type of rapid acquisition, the Commercial Solutions Opening (CSO). The CSO is a solicitation technique that is designed as an innovative means to solve the problem of slow government procurement. At its core, the CSO seeks to take a broadly identified objective, stated in a manner that allows for diverse solutions, and award a contract to meet those objectives within a matter of weeks, as opposed to the methods that now take months or even years using "traditional" models. A CSO can result in both FAR-based and non–FAR-based contracts and is used to acquire an innovative technology or an innovative means or method to accomplish the objective.

While innovation is specifically defined in the FY2022 NDAA (2021) as "(1) any technology, process, or method, including research and development, that is new as of the date of submission of a proposal," or "(2) any application that is new as of the date of submission of a proposal of a technology, process, or method existing as of such date" (p. 275), innovation does not require the solution be completely new or never-before attempted. In fact, the CSO community even refers to simple maintenance activities like grounds maintenance as candidates for CSOs, if the agency seeks an innovative means or method of achieving these common tasks (82d Contracting Squadron, 2020).

For the many flexibilities and efficiencies that a CSO provides, it is important to also recognize how not to use a CSO. Based on our collective research from various sources and experiences, a CSO is not a solicitation technique to obtain services where the government already has the requirement defined, a solicitation technique to obtain standard technological configurations or support where the government has a design specification, a solicitation technique to shortcut competition or except fair opportunity, or a quick sourcing solution for poorly defined requirements (Secretary of the Air Force Acquisition Office [SAF/AQCP], 2022). The next section describes the purpose of this research.

The primary purpose of this research is to identify the strengths, weaknesses, and best practices of the CSO as a solicitation technique leading to a contract award. This research intends to provide DoD organizations and their workforces with a consolidated report analyzing available data on the CSO solicitation technique and making recommendations based on the use of CSOs. Following the purpose of the research, the next section specifies the research questions with which we hope to achieve the purpose.

Research Questions

This research is framed by the understanding that traditional FAR techniques can be ineffective at acquiring innovative solutions (GAO, 2019). This research explores opportunities and flexibilities of CSOs as a solicitation technique to acquire innovative solutions and seeks to answer the following questions:

1. What are CSOs' strengths as a solicitation technique?

2. What are CSOs' weaknesses as a solicitation technique?

3. What are best practices for utilizing the CSO solicitation process?

4. What is the statistical difference, if any, in the procurement lead times of contracts awarded from a CSO and those awarded from a FAR-based solicitation, and what inferences can be made of this difference?

Methodology

This research assesses the strengths, weaknesses, and best practices of CSOs as a solicitation technique in acquiring innovative solutions. Extensive direct feedback will be captured from CSO Cross Talk meetings among DoD agency points of contact who have previously conducted CSOs and/or are working to develop CSO policies/procedures at their individual agencies. These feedback meetings are led by the Secretary of the Air Force Acquisition Office (Contracting; SAF/AQC). This information is reviewed for strengths and/or weaknesses regarding training and information sharing, internal agency processes, solicitation definition, and industry interaction. Assessment of different individuals' varied interpretation and implementation of the flexible process to meet their specific program and agency goals informs the categorization of strengths, weaknesses, and best practices. Similar direct user feedback is discussed as compiled for and documented in other published briefings and reports. The research also attempts to quantify DoD's procurement lead time using data from the FPDS and determines if there is a statistical difference in the procurement lead time of contracts awarded from a CSO and those using a FAR-based solicitation. The results are presented in the form of recommendations that the DoD and its contracting offices can use to best implement CSOs. Following the research methodology, the next section provides the intended and anticipated benefits of this research.

Innovation Theory and Commercial Solutions Openings

Innovation in business is the foundation for examining CSOs and their benefits. CSOs present an opportunity for the DoD to make critical investments in technology and capability by leveraging the technological capabilities of the department's industrial base. In fact, the adoption of CSOs as a permanent authority is itself, innovative. To understand how these innovative capabilities can shape the DoD, it is important to understand the theory supporting innovation in business, including the different paradigms that are found in literature. First, we must consider the DoD as a type of knowledge management (KM) firm with "roles and processes to support decision-making" (Neary, 2018. p. 1). The DoD as a KM firm is comprised of individuals with tacit, explicit, and implicit knowledge of the military's operations, from munitions flight trajectories to the ideal length of a blade of grass along a flightline. Within this construct, the DoD is operating as a firm competing with other nations; this defines the marketplace within which innovation leads to competitive advantage and provides a framework against which innovation theory can be applied.

Deeper View of Research Methodology

With strengths, weaknesses, and best practices at the core of this research's primary questions, it is important to define those terms. A *strength* indicates an aspect of the CSO solicitation technique that has benefited the government, industry, or both. Examples could include an easier process to contract award than FAR-based procedures, reduced risk of protest, contracts for more innovative solutions than the government could have defined in a requirements statement, and so forth. A *weakness* would indicate an aspect of the CSO that has hindered the government, industry, or both. Examples could include a more confusing process than FAR-based procedures, difficulty in securing a fair and reasonable price for the government, uncertainty for how to award follow-on contracts to initially innovative solution contracts, and so forth. An observation may have attributes that result in both a strength and weakness.

A *best practice* is defined by Merriam-Webster (n.d.) as “a procedure that has been shown by research and experience to produce optimal results and that is established or proposed as a standard suitable for widespread adoption.” Examples could include implementing an agency-specific CSO guidebook, using a gated/phased approach for CSO proposal submissions, advertising CSOs through unconventional means, and so forth. Not all observations may qualify as a strength, weakness, or best practice but still enhance or contribute to this research or areas for future research; those observations are captured as “other observations” later in this paper. The next section describes the methodology for gathering CSO Cross Talk data.

Commercial Solutions Opening Cross Talks

CSO Cross Talk meetings started being held quarterly in April 2022 as a forum for the DoD contracting workforce to share “CSO policy changes, training, and success stories/best practices” (DoD, 2022). DoD agency points of contact who have previously conducted CSOs share varied interpretation and implementation of the flexible solicitation technique to meet their specific program and agency goals. This is to benefit all those working to develop CSO policies/procedures at their individual agencies, whether they have used them yet or not. Participants are encouraged to ask questions and suggest hot topics surrounding CSOs. SAF/AQC representatives organize and facilitate the meetings, and afterwards, they draft CSO Cross Talk Bulletins to summarize the meetings. These bulletins are disseminated with guidance for meeting attendees to share among their respective DoD agencies’ acquisition workforces.

For this research, the contents of these bulletins, primarily based on the feedback provided by DoD agency points of contact who have previously conducted CSOs, will be reviewed and analyzed, particularly regarding CSO strengths, weaknesses, and best practices. While a policy analyst or contracting officer may just read these bulletins and try to take mental notes for potential future use, this research will systematically break down all feedback data and categorize it by topic area to lend itself more readily to making strategic recommendations about actions that can be taken regarding CSOs. The four overarching categories are

- **training and information sharing:** how the workforce is educated on this solicitation technique,
- **internal agency processes:** how individual DoD agencies structure their facilitation of evaluating and awarding CSOs,
- **solicitation definition:** how various contracting officers draft individual CSOs, and
- **industry interaction:** how the government advertises to and receives information from potential offerors.

These four categories are purposely broad to accommodate finding space for a diverse range of feedback since the DoD agency points of contact were not required to structure their Cross Talk presentations in any way. Once the feedback is separated into these categories, then strengths, weaknesses, and best practices can be identified among them. Further, commonalities and focus areas for recommendations can be consolidated. The next section discusses the research methodologies to be used in analyzing other published briefings and reports.

Other Published Briefings and Reports

Published briefings and reports are reviewed from various sources including congressional committees and GAO reports. The contents of the reports are analyzed for strengths, weaknesses, and best practices, and then categorized accordingly. The GAO and U.S. Court of Federal Claims (COFC) archives are also reviewed for protest reports. The contents of these reports are analyzed for strengths, weaknesses, and best practices, and then categorized accordingly. The next section discusses the methodologies to analyze data about CSOs and resulting contract awards.

Procurement Lead Time Data Analysis

The Defense Innovative Unit has realized notable decreases in their acquisition time lines by using CSOs. This research attempts to quantify DoD's procurement lead time efficiencies using data from the FPDS. The FPDS is a data reporting tool that captures contract data about each reportable contract action, that is each contract action over the micro-purchase threshold, including awards, modifications, and orders (FAR 4.6, 2023). Data are then made available through the System for Award Management (SAM) reporting tools and can be analyzed across a myriad of data fields. SAM reports can produce standard reports containing predefined criteria, or a user can create ad hoc reports within which the user can define the specific criteria, including filters, reported fields, and format. To support reporting, the General Services Administration (GSA) maintains a Data Element Dictionary that explains each available data element collected through contract action reporting (GSA, 2023). This research uses ad hoc reports of contract data with the report criteria as provided in Table 1.

Table 1. SAM Ad Hoc Report Criteria (GSA, 2023)

Field	Description[a]	Criteria
Date Signed	"The date that a mutually binding agreement was reached" (p. 23)	Oct 1, 2019 ≤ date signed ≤ Jan 1 2023
Solicitation Date	The date the solicitation was issued	Oct 1, 2019 ≤ solicitation date
Base and All Options Value (Total Contract Value)	"The mutually agreed upon total contract value including all options (if any)" (p. 30)	<$100,000,000
Contracting Agency ID	"The code for the agency of the contracting office that executed or is otherwise responsible for the transaction" (p. 37)	Equals 1700 (Navy), 2100 (Army), and 5700 (Air Force)
Solicitation ID	"Identifier used to link transactions in FPDS to solicitation information" (p. 20)	Is Not Null
Modification Number	"An identifier … that uniquely identifies one modification for one contract, agreement, order, etc." (p. 17)	Equals 0

We conducted two-sample *t*-test analyses of procurement times for each population set. Through the analyses we attempted to determine whether a significant difference in procurement times exists between the CSO solicitation process and the FAR-based solicitation approach. Procurement time was defined as the days from the solicitation issuance date to the date of award, comparing mean procurement times for acquisitions that use CSOs with that of requirements sourced through FAR-based means such as requests for quotes and requests for proposals. As multiple awards can be made from a single CSO, only the days-to–first award were considered. Days-to–first order were determined by considering the total set of awards issued pursuant to a CSO solicitation and selecting the earliest award date to include in the CSO sample. Only FAR-based awards made between October 1, 2019, and January 31, 2023, were considered. Data were segregated into eight distinct populations in sets of two, resulting in one population set for actions below the Simplified Acquisition Threshold (SAT), one population set for actions between the SAT and $4.99 million, one population set for actions between $5 million and $99.99 million, and one population set for all actions below $100 million. CSOs were identified by the inclusion of "S" and "C" in the ninth and 10th positions of the solicitation ID, allowing for the segregation of the data into the two distinct population sets.

Table 2. Description of Populations and Notations for Statistical Analysis

Population Set	Population	Criteria	Notation Example
(i) Below SAT		Contracts with award value < $250,000	
	Awards from CSO solicitation		$N_{CSO(i)}$
	Awards from FAR solicitation		$N_{FAR(i)}$
(ii) Between SAT and $5 Million		Contracts with award value ≥ $250,000 and < $5 million	
	Awards from CSO solicitation		$N_{CSO(ii)}$
	Awards from FAR solicitation		$N_{FAR(ii)}$
(iii) Above $5 Million		Contracts with award value ≥ $5 million and < $100 million	
	Awards from CSO solicitation		$N_{CSO(iii)}$
	Awards from FAR solicitation		$N_{FAR(iii)}$
(iv) Total Population		All contracts with award value <$100 million	
	Awards from CSO solicitation		$N_{CSO(iv)}$
	Awards from FAR solicitation		$N_{FAR(iv)}$

Table 3. Description of Population Justifications

Population Set	Justification
(i) Below SAT	Acquisitions under the SAT are generally expedited when compared to non-SAT acquisitions, regardless of the solicitation methodology chosen; therefore, the SAT provides a logical cutoff for the first set population set.

(ii) Between SAT and $5 Million	Acquisitions of $5 million and above have additional reviews and approvals required by many agencies. For example, the Air Force, which has executed the preponderance of DoD's CSOs, requires additional clearance reviews starting at $5 million. To ensure parity in the data, $5 million is used as the demarcation point to segregate the data samples.
(iii) Above $5 Million	CSOs above $100 million require special approval from the Under Secretary of Defense for Acquisition and Sustainment (USD[A&S]); therefore, the procurement time is elongated through additional reviews and oversight. There are many additional factors for these larger-dollar procurements that challenge comparison with the data presently available through SAM. Analysis of actions above $100 million requires a level of analysis that exceeds the scope of this research; therefore, these actions are excluded from the statistical analysis.
(iv) Total Population	The total population sets of CSOs and FAR solicitations resulting in award below $100 million, enabling a wholistic analysis of the two distinct populations.

Collectively, each population set was tested against the following hypothesis with a confidence interval of CI = .90 (α = .10).

H_0: $\mu_{CSO} = \mu_{FAR}$ (1)

H_1: $\mu_{CSO} \neq \mu_{FAR}$(2)

As discussed earlier, data quality and quantity are limitations of this research. The quantity of CSO data may not be sufficient to test the hypothesis for each population set; in those instances we made informed inferences from the available data. Further, the quality of FPDS data may necessitate the elimination of outliers from the data sets; in the event outliers are removed, they are addressed in the capstone project in detail.

Once the *t*-test analysis is complete, it may be possible to further subdivide the data into individual agencies to aid future research.

Author's note: Please see the NPS capstone project for a complete description of methodology, framework for analysis, results, and findings.

Implications of Findings

Most of the listed CSO Cross Talk comments were categorized as best practices since the agency representatives primarily framed their feedback as subjective recommendations to other agencies. Objective strengths and weaknesses may have been few because of the noted lack of accurate CSO data reporting. It is possible to infer that some of the best practices could be due to a strength being the flexibility of the CSO solicitation technique. Alternatively, a weakness being ambiguity or confusion could also be inferred when considering the extensive best practices, with the majority regarding Internal Agency Processes, being recommended to ensure efficiency and successful contracts, which may otherwise not be achieved. The most comments being categorized under Internal Agency Processes is also notable in the types of recommendations that the acquisition community feels are needed and will be well-received and utilized. Finally, it is noted that a few of the observations are duplicative, but they were all left in to highlight how multiple agencies made similar comments as that could influence prioritization of recommendations at the end of this paper. Expanding beyond just the limited number of strengths and weaknesses identified in the CSO Cross Talk feedback, the other findings discussed in this paper capture that there are overall many more strengths than weaknesses regarding CSOs at this point.

In total, we made 66 individual observations of strengths, weaknesses, and best practices. Within those observations we identified 27 strengths, seven weaknesses, and 43 best practices in the documented findings of the CSO data. Some of these observations were assigned to multiple categories or were defined as both a best practice and a strength or a weakness. These findings were also categorized across 10 categories according to their central theme(s), with some findings falling into multiple categories. The total quantities of strengths and weaknesses by category were captured. The protest findings, especially, are a very telling representation of the significant advantage that CSOs may have over FAR-based solicitation techniques in that so few protests have been filed, and none have been sustained that were filed based on the CSO process itself. Additionally, the process flexibility and limited scope of litigation that comes from judicial deference are strengths that merit prudent planning and potential opportunities which contracting activities can embrace in their own solicitation planning process.

Category	Strengths	Weaknesses
Training and Information Sharing	1	1
Internal Agency Processes	2	4
Solicitation Definition	2	0
Industry Interaction	1	0
Expanded Solution Horizons	4	0
Industry Participation and Competition	3	0
Cost/Price/Budgeting	1	1
Schedule and Planning	1	1
Process Flexibility	7	0
Scope of Litigation	5	0

Figure 1. Quantity of Strengths and Weaknesses by Category

The procurement lead time analysis results are provided in Table 3. Upon reviewing these results, one may surmise that the CSO solicitation process is wholly inefficient at expediting the time to contract award; however, this analysis is a singular facet of the total research and is constrained by factors that preclude definitive decision-making regarding the procurement lead time. Regardless, the procurement lead time analysis does not support that CSOs are an expedited acquisition technique. The analysis of procurement lead time discussed in this is constrained by the quality and quantity of the available data. For this research, we performed a statistical analysis of the CSO procurement lead time by quantifying the days that elapsed from the CSO issue date to the earliest date of contract award made from the CSO. This analysis relied on the data input to FPDS by contracting activities reporting contract awards. While we recognize that some CSO models allow for initial responses to be received many days or even months after the CSO is issued, the data available in FPDS does not provide for a means to identify the elapsed time between CSO responses and contract award. Further, the solicitation date is manually entered into the system by the contracting activity, leaving room for user error and misreporting. These factors exemplify the quality and quantity constraints identified in this research and do not provide for an infallible method of testing the CSO process as compared to the FAR solicitation techniques. Even so, our procurement lead time analysis provides for a foundational baseline and analytical model against which future analysis may be conducted, once more reliable data can be obtained through implementing the recommendations discussed in the next section. With improved data quality and reliability, the model we established in this research will facilitate a more robust and reliable comparison of the CSO process and FAR solicitation techniques, allowing for validation, verification, and representative quantification of the strengths and weaknesses identified in this research.

Table 4. Procurement Lead Time Analysis Results (Statkat, n.d.)

Population Set	*df*	*t*	*t*-critical[a]
(i) Below SAT	6.020	.294	1.942
(ii) Between SAT and $5M	17.020	1.701	1.739
(iii) Above $5M	5.006	.872	2.015
(iv) Total Population	30.015	2.412	1.697

[a]Retrieved from statkat.com, (Online calculator, n.d.)

Given the totality of the research we have conducted, we believe that the CSO process should be embraced by agencies seeking to expand their technological horizons and capabilities. The strengths we identified in this research greatly outweigh the weaknesses. Using the best practices and observations we have noted in our research, agencies can equip themselves with the best means and processes to execute successful CSO solicitations. From the data, we find that the CSO solicitation technique also has applications beyond the research and development arenas and can be used to identify innovative means to accomplish operations, sustainment, and even maintenance tasks, potentially providing total life-cycle cost savings to the government as a result. As discussed throughout this research, we also note that the CSO process and procedure is relatively immature and rapidly evolving as compared to other solicitation methodologies. To ensure the continued success of the CSO as a solicitation technique to achieve innovation, we provide targeted recommendations in the areas of training and development, policy changes, and tracking and reporting, which are contained in the next section of this research.

Recommendations

This section presents focused recommendations based on the results of the analysis found in this research. In total we provide eight recommendations, each with their anticipated benefits and methods to implement. The recommendations encompass three categories: training and development, policy changes, and tracking and reporting.

Federal Procurement Data System Modification

The first recommendation involves both a policy change and a tracking and reporting change. We perceive this recommendation to be the simplest to implement. FPDS data are collected through contract action reporting. This reporting is completed by individual contracting activities completing a form in the system, which provides data about the contract(s) reported. To meet the government’s reporting needs and requirements of the time, these form fields are often updated and changed, and new fields are added as necessary. This includes the addition of new data elements, new reporting options, and temporary instructions through special coding in the description field. These changes are executed by a team of support contractors.

We propose a two-part modification to the FPDS contract action report. The first modification is to include Solicitation Technique as a reporting criterion. This field would capture the solicitation technique used to acquire the contract award being reported and should include a drop-down selection for CSO as well as ones for other solicitation techniques such as request for proposal, request for quote, BAA, invitation for bid, and others. With the addition of the Solicitation Technique reporting criterion, the government and future researchers will be able to analyze specifics about solicitation methodologies and the contract awards that follow in a manner like the analysis we conducted in this research. The inclusion of the Solicitation Technique reporting criterion will also allow for the analysis of other areas that extend beyond the scope of our research, such as industry involvement across differing solicitation techniques, cost/price history and modification metrics, small business participation across solicitation techniques, and targeted areas to bolster training in solicitation techniques. Absent a dedicated field to report solicitation technique, we recommend the government modifies the action description field to enable reporting of the solicitation technique, which would still present opportunities for future reporting, analysis, and informed decision-making.

The second modification to the FPDS contract action report we recommend is the inclusion of Initial Proposal Receipt Date as a reporting criterion. This new field should be a date field that reports the date the initial proposal was received for all new awards being reported into the FPDS. The FPDS contract action report currently includes a field to report the solicitation date; however, this is not necessarily a useful data point for general solicitations, which can be open for long periods of time and which can invite multiple proposals during its open period(s). Absent this modification to the FPDS, there is no discernable means to distinguish the procurement lead times between a contract action where the proposal was received 1 day after the CSO was issued, and a contract action where the proposal was received 1 year after the CSO was issued. The addition of proposal receipt reporting will enable future analysis of procurement lead time for both contracts awarded from CSO solicitations, and those awarded by other means.

Expand Contract Type Options

The next policy change recommendation involves a more material revision to the CSO authority by expanding the available contract types for awards to include time and materials or labor hour. Since CSOs are soliciting innovative solutions, it is reasonable to assume that offerors may not always be able to precisely estimate the work required to achieve their potentially groundbreaking goal. It would be doing a disservice to the government to lose the possibility of awarding a contract for that product, technology, or service because the offeror did not want to submit a fixed price proposal and risk its profit potential if it took more effort or resources to complete the contract objectives than the contractor had proposed. This recommendation could be considered by Congress to expand the language of Section 803 of the FY2022 NDAA (2022) to include provisions of expanded contract types in awards from CSOs. The Office of Defense Pricing and Contracting could then issue a new class deviation recognizing the expanded authority. While this research only considered data and literature available as of January 31, 2023, it is noted that on that day, DoD proposed amendments to the Defense Federal Acquisition Regulation Supplement (DFARS) to add the preponderance of Class Deviation 2022-O0007 into DFARS Part 212, with public comments due April 3, 2023 (Defense Federal Acquisition Regulation Supplement [DFARS] PGI, 2023).

We note that the DFARS has been revised. Class Deviation 2022-O0007 no longer provides the authority for CSOs. CSOs are now included at DFARS 212.70 with slightly different language (e.g., funds availability is no longer required to be a primary evaluation factor).

Formal Training through the Defense Acquisition University

For the first training and development recommendation, we recommend the Defense Acquisition University (DAU) develop and offer a standalone training course on CSOs. It should begin with comparing the differences from FAR-based solicitation techniques and identifying the processes and/or documentation that it bypasses for the special purpose of streamlining contract awards for innovative solutions, like how we have conducted our research. Our research and findings can even be used as a starting point to develop the course material, or our research could be included in its entirety to facilitate critical thinking and analysis through the DAU course. Since there are so many different uses under the CSO authority's definition of "innovative," it would be prudent for more contracting officers to have the opportunity to learn about the authority and its opportunities, add it to their contracting toolbox, and champion for its implementation when possible and appropriate at their individual agencies. The course can also provide its students with solicitation and evaluation templates and plain language documentation to use as a resource. As highlighted often in the CSO Cross Talks, while CSO flexibility is appreciated, there is great value in standardization and uniformity for repeatable processes. As a future evolution of this training and development recommendation, the DAU, or some other activity, could develop a comparative tool that includes decision logic to guide future procurement teams through a methodical decision process of choosing the most advantageous solicitation technique for their requirement(s), whether that be a CSO or some other solicitation technique.

Invest in Commercial Solutions Opening Center of Excellence

This recommendation expands upon the original recommendation by Washburn and Colavito (2023) and recognizes that the U.S. Air Force (USAF) has adopted Washburn and Colavito's original recommendation to "Establish Commercial Solutions Opening Center of Excellence." We recommend that the USAF fully invest in the CSO Center of Excellence and take the DoD lead in consolidating CSO DoD guidance documents, best practices, and procedures in furtherance of the DoD's KM environment. These resources could be documented and catalogued through a virtual site with appropriate access controls, perhaps as a resource open to all DoD access card holders under the USAF Innovation Toolbox (United States Air Force, n.d.). A similar website after which to model itself could be the "Acquisition Innovation" site created and maintained by the Defense Advanced Research Projects Agency, which features history, training, samples, and other resources for the acquisition of innovative technology using the award of OTs (Defense Advanced Research Projects Agency, n.d.). As the CSO Center of Excellence, the USAF should maintain flexibility in remaining current with best practices regularly being discovered and shared as more CSOs are being utilized. The CSO Center of Excellence should also explore opportunities to develop meaningful data analytics and metrics to measure CSO utilization and effectiveness as resulting contracts are performed. Furthermore, the CSO Cross Talks should be continued for which policy advisors and experienced practitioners can still directly contribute, but their resultant summary bulletins and other guides, samples, and so forth can be shared for any DoD acquisition personnel on the recommended virtual site.

Addressing Resource Strain through Organizational Structuring

Beyond the individual contracting officer training and development, a key recommendation is for senior contracting officials to recognize the resource strain that may result in the use of CSOs and to develop organizational structures accordingly. While the CSO is touted as an easy and streamlined process, it has been anecdotally proven in the CSO Cross

Talks and our own observations to become administratively cumbersome to manage when there is a high likelihood of strong interest from industry to submit proposals. Depending on the agency's structure, separate CSO divisions and additional personnel may be necessary to ensure the potential efficiencies can be maximized. Contracting offices must also ensure they achieve buy-in from their agency's technical subject matter experts and all necessary agency stakeholders, such as information technology, cybersecurity, and logistics, to facilitate prompt proposal review, operational feasibility, and close collaboration with the contracting officer(s) to draft successful contracts.

Publication of Requirements and Industry Involvement

Another recommendation is regarding industry engagement as numerous findings point to the need for creative means to interact with potential offerors. To successfully reach the often-nontraditional companies that may otherwise be intimidated or discouraged by FAR-based solicitation techniques, DoD agencies need to make effort to advertise their CSOs beyond the GPE. Links to the CSO posted on LinkedIn or industry-specific websites would be helpful. Beyond that, technical subject matter experts or contracting personnel could attend industry conferences to have one-on-one networking opportunities with the types of companies they think could have government-applicable innovative ideas. This recommendation can be categorized under training and development as it deviates from traditional solicitation publication methods, and the acquisition workforce will need education on the value of taking these extra steps beyond the usual process. As discussed previously, the posting of the CSO mimics a combination of market research techniques and the solicitation; embracing this recommendation takes advantage of this opportunity for efficiency and evolves it through combining additional pre-award elements of information sharing (FAR 5.1, 2023), leading to further opportunities for efficiency.

Improve Reporting of Negotiation Documentation to Capitalize on the Department's System of Systems

Our penultimate recommendation addresses a final policy, tracking, and reporting change. When conducting negotiations of noncompetitive contract actions valued above $25 million, contracting officers are required to upload approved negotiation documents, such as price negotiation memorandums, into the Contract Business Analysis Repository (CBAR) tool in the government's Procurement Integrated Enterprise Environment (PIEE) suite of applications (DFARS PGI, 2023). The results of the negotiations are then made available to other contracting personnel to prepare for future negotiations. Further, when uploading the negotiation documents, users are required to enter basic information about the agency, contractor, contract, and negotiation process. Unfortunately, to retrieve details about the negotiation and reasonableness determination process(es), users must scour the tool, download, and read through negotiation documents individually to understand the negotiation history. As part of the PIEE suite, the CBAR tool connects to the Electronic Document Access application, which provides for post-award administrative reporting. CBAR could also connect to other applications and tools within the PIEE to form a system of systems and enable robust reporting and business analytics.

Considering CBAR's utility as a tool to assist future negotiations, and in acknowledgement of the CSO process, which is considered competitive, we first recommend a policy change that expands the mandatory reporting requirement and upload of cost/price negotiation documents for *all* contract actions valued above $25 million *regardless of the competitive nature of the requirement*. The requirement to determine a price fair and reasonable is universal and does not distinguish between whether the action is competitive or noncompetitive. Our recommendation recognizes that when negotiations occur, FAR 15.406-3 requires that those negotiations be documented in some form. CSOs are not exempt from this documentation requirement when the contracting officer engages in negotiations. This change will provide additional resources to contracting officers in developing future negotiation objectives for both CSOs and those using FAR-based techniques by expanding the pool of available resources useful for preparing for and establishing negotiation objectives.

Expanding the reporting requirements does not address the accessibility flaw of the CBAR tool. Acknowledging the scalability of the PIEE suite, we further recommend the CBAR tool be modified to include a field that requires solicitation and evaluation methodology when uploading a negotiation document. Including this field will enable a more streamlined method to conduct reviews and analyses of how fair and reasonable pricing is achieved for both CSOs and all other reportable contract actions. Further, even for contracts that do not exceed the minimum reporting threshold established in the DFARS PGI, the DoD should consider requiring reporting of the process(es) used to determine fair and reasonable pricing, especially for commercial acquisitions, including those that used the CSO solicitation technique. This requirement will provide an array of valuable data, bolstering the negotiating process and lessening the narrow reliance on business acumen to determine price reasonableness. Scaling the CBAR tool could then lead to further applications to support negotiations, such as connection points with the USAF's weighted guidelines online tool and others, but those applications are beyond the scope of this research and its recommendations.

Caution Against Wide-Sweeping Changes in Policy

As a final recommendation, we recommend constraining future policy regarding the CSO solicitation technique to only those necessary to execute legal contracts and agreements. As reflected in this research, innovation requires flexibility and freedom to engage in continuous improvements and limit imitation. To maintain the flexibility of CSOs, future policy should avoid unnecessary restrictions in the CSO process. Rather than policy that constrains or restricts the CSO solicitation process, the government should instead invest in its KM environment and bolster the government workforce's knowledge and understanding of CSOs to facilitate further innovation in the procurement process. Doing so will equip the DoD workforce with the "best weapons with which to compete … knowledge and service" (Johannessen et al., 1999, p.132). This will lead to increased learning capacity of the DoD's knowledge workers and secure a competitive advantage of defense superiority. The CSO process and this recommendation, taken collectively with our other recommendations, will facilitate the DoD securing this competitive advantage through KM.

Summary and Conclusion

The goal of this research was to answer four research questions. The questions were intended to explore the opportunities and flexibilities of CSOs as a solicitation technique to acquire innovative solutions. This information could then be used to frame DoD agencies' utilization of the CSOs to support their individual missions. While not definitively answered due to limitations in the research, the following conclusions to the research questions have been made based on our findings:

1. What are CSOs' strengths as a solicitation technique?

Through this research, we identified 27 strengths of the CSO process. These fell across 10 distinct categories:

- training and information sharing (number of findings = 1),
- internal agency processes (2),
- solicitation definition (2),

- industry interaction (1),
- expanded solution horizons (4),
- industry participation and competition (3),
- cost/price/budgeting (1),
- schedule and planning (1),
- process flexibility (7), and
- scope of litigation (5).

Some strengths were assigned to multiple categories. The most telling and compelling strengths were identified in the GAO and COFC protest findings, namely that CSOs may have a significant protest-risk advantage over FAR-based solicitation techniques as there have been zero sustained protests that challenged the CSO process itself. Additionally, the judicial deference provided to the CSO process by GAO and COFC appreciably enhance the protest-risk advantage of using CSOs to acquire innovation.

2. What are CSOs' weaknesses as a solicitation technique?

Through this research, we identified seven weaknesses of the CSO process. These fell across four distinct categories:

- training and information sharing (number of findings = 1),
- internal agency processes (4),
- cost/price/budgeting (1),
- schedule and planning (1).

Though few compared to the total strengths, the CSO weaknesses point to the need to engage in prudent planning and develop sound processes when planning a CSO solicitation. Particularly, we find the absence of weaknesses identified in the GAO and COFC protest decisions to be noteworthy.

3. What are best ractices for utilizing the CSO solicitation process?

Through this research, we identified 43 individual best practices for implementing the CSO process. These best practices involved the planning process, the soliciting process, and the evaluation process. We recommend adoption of the entire catalogue of best practices when planning future CSO solicitations. Of note are the best practices regarding internal agency processes, as this category had the most robust list of recommendations from early CSO users. As CSOs become a more popular solicitation technique for both the government to use and industry to respond to, agencies will need to recognize the importance of properly scaling up in their preparation of the planning, soliciting, and evaluation processes surrounding it.

4. What is the statistical difference, if any, in the procurement lead times of contracts awarded from a CSO and those awarded from a FAR-based solicitation, and what inferences can be made of this difference?

Taken individually, the procurement lead time data analysis suggests that no significant difference exists between the procurement lead times of contracts awarded from a CSO and those using a FAR-based solicitation when examining them in three distinct groups of (1) less than the SAT, (2) SAT to less than $5 million, and (3) $5 million to $100 million. Considering the data as a collective of all actions less than $100 million, however, the analysis found that a statistical difference does exist in the procurement lead times of contracts awarded from a CSO and those using FAR-based solicitations. This finding, in conjunction with the finding that the mean procurement lead time of contracts awarded from a CSO is longer than the mean procurement lead time of those contracts awarded using a FAR-based solicitation, suggests that the procurement lead time for contracts awarded from a CSO is significantly longer than those using FAR-based methodologies when considering the totality of all actions less than $100 million.

Considering these findings, one might surmise that the CSO process is wholly inefficient at expediting the time to contract award; however, this analysis is a singular facet of the total research and is constrained by factors that preclude informed decision-making regarding the procurement lead time. As discussed previously, our statistical analysis of procurement lead time is constrained by the quality and quantity of the available data. Due to these constraints, we were unable to make reliable, informed inferences about the procurement lead times; however, we postulate that our analysis provides for a foundational baseline and analytical model against which future analysis may be conducted, once more reliable data can be obtained through implementing our recommendations contained in this research. With improved data quality and reliability, the model we established in this research will facilitate a more robust and reliable comparison of the CSO process and FAR solicitation techniques, allowing for validation, verification, and representative quantification of the strengths and weaknesses identified in the research.

Call to Action

In summary, CSOs provide an opportunity for the DoD to capitalize on the innovative capabilities and advances of industry, propelling the DoD to expanded solutions horizons, improving industry participation and competition, providing process flexibility, and securing against protest risk. As a solicitation technique, the CSO is a valuable tool to achieve innovation, but prudent planning and application of this research's identified best practices are critical to ensure acquisition success. Further, by implementing the recommendations provided in this research, the DoD will be postured to utilize the CSO solicitation technique to its fullest potential, closing the technological capability gap and providing better defense capabilities to the nation.

References

82d Contracting Squadron. (2020, June 3). *AF customer education program - CSO: Overcoming barriers* [Video]. https://www.milsuite.mil/video/watch/video/31919

Defense Advanced Research Projects Agency. (n.d.). *Acquisition innovation*. Retrieved May 7, 2023, from https://acquisitioninnovation.darpa.mil/

Defense Federal Acquisition Regulation Supplement PGI, 48 C.F.R. 215 (2023, May 4).

DoD. (2022, April 27). DOD quarterly commercial solutions opening cross talk (Issue 1) [Bulletin].

DoD Inspector General. (2015). *Inspector General's summary of management and performance challenges for Fiscal Year (FY) 2015*. https://media.defense.gov/2017/Apr/14/2001733120/-1/-1/1/FY2015_DODIG_SUMMARYMANAGEMENT_CHALLENGES.PDF

DoD Inspector General. (2022). *Fiscal Year 2023 top DOD management challenges*. https://media.defense.gov/2022/Nov/16/2003115791/-1/-1/1/MANAGEMENT%20CHALLENGES%20FY2023.PDF

Federal Acquisition Regulation, 48 C.F.R. 4 (2023, January 31).

Federal Acquisition Regulation, 48 C.F.R. 5 (2023, May 8).

GAO. (2019). *DOD acquisition reform: Leadership attention needed to effectively implement changes to acquisition oversight*. https://www.gao.gov/assets/gao-19-439.pdf

GSA. (2023, January 28). *GSA Federal Procurement Data System (FPDS) data element dictionary*. https://www.fpds.gov/downloads/Version_1.5_specs/FPDS_DataDictionary_V1.5.pdf

Johannessen, J.-A., Olsen, B., & Olaisen, J. (1999). Aspects of innovation theory based on knowledge-management. *International Journal of Information Management*, 121–139.

Merriam-Webster. (n.d.). Best practice. In *Merriam-Webster dictionary*. Retrieved April 15, 2023, from https://www.merriam-webster.com/dictionary/best%20practice

National Defense Authorization Act for Fiscal Year 2016, Pub. L. No. 114-92 § 804 (2015). https://www.congress.gov/114/plaws/publ92/PLAW-114publ92.pdf

National Defense Authorization Act for Fiscal Year 2017, Pub. L. No. 114-328 § 879 (2016). https://www.congress.gov/114/plaws/publ328/PLAW-114publ328.pdf

National Defense Authorization Act for Fiscal Year 2017, Pub. L. No. 114-328 § 880 (2016). https://www.congress.gov/114/plaws/publ328/PLAW-114publ328.pdf

National Defense Authorization Act for Fiscal Year 2018, Pub. L. No. 115-91 § 867 (2017). https://www.congress.gov/115/plaws/publ91/PLAW-115publ91.pdf

National Defense Authorization Act for Fiscal Year 2020, Pub. L. No. 116-92 § 861 (2019). https://www.congress.gov/116/plaws/publ92/PLAW-116publ92.pdf

National Defense Authorization Act for Fiscal Year 2022, Pub. L. No. 117-81 § 803 (2021). https://www.congress.gov/117/plaws/publ81/PLAW-117publ81.pdf

Neary, S. (2018). *Knowledge and information management*. Joint Staff J7. https://www.jcs.mil/Portals/36/Documents/Doctrine/fp/knowledge_and_info_fp.pdf?ver=2018-05-17-102808-507

Secretary of the Air Force Acquisition Office (Contracting Policy / Regulation). (2022, November 2). *Is a CSO right for you* [Video]. https://usaf.dps.mil/sites/AFCC/KnowledgeCenter/Documents/Forms/AllItems.aspx?id=%2Fsites%2FAFCC%2FKnowledgeCenter%2FDocuments%2FCSO%5FLibrary%2FIs%20a%20CSO%20Right%20for%20You%2Emp4&parent=%2Fsites%2FAFCC%2FKnowledgeCenter%2FDocuments%2FCSO%5FLibrary&p=true&ga=1

Section 809 Panel. (2018). *Report of the Advisory Panel on Streamlining and Codifying Acquisition Regulations*. https://discover.dtic.mil/wp-content/uploads/809-Panel-2019/Volume1/Sec809Panel_Vol1-Report_Jan2018.pdf

Statkat. (n.d.). *Critical t value given significance level alpha—Online calculator*. Retrieved from https://statkat.com/online-calculators/critical-t-value-given-alpha.php

United States Air Force. (n.d.). *Innovation toolbox*. Retrieved May 6, 2023, from https://usaf.dps.mil/sites/AFCC/AQCP/OI/Innovation%20Toolbox/SitePages/Solicitation.aspx

Washburn, E., & Colavito, M. (2023). *Analysis of Department of Defense commercial solutions openings* [Capstone applied project report, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/72289

PANEL 17. PLANNING FOR AUTONOMOUS VESSELS

Thursday, May 9, 2024	
10:30 a.m. – 11:45 a.m.	**Chair: Capt. Dennis Monagle, USN,** Program Manager Multi-Mission Tactical Unmanned Aerial Systems (UAS) ***Unmanned Low Profile Vessels (ULPVs): "Narco Subs" for Contested Logistics*** Sergio Sierra, Naval Postgraduate School ***System Product Line Cost and Investment Modeling Applied to UUVs*** Raymond Madachy, Naval Postgraduate School ***Next Generation Logistics Ship Automation and Uncrewed Underway Replenishment*** Nabil Tahan, Naval Postgraduate School

Capt. Dennis Monagle, USN—is the program manager Multi-Mission Tactical Unmanned Aerial Systems program office (PMA-266) in March 2022. He is a native of Athens, Georgi and graduated from Indiana University in 1994 with a bachelor's degree in applied science. He earned his master's degree in management from the University of Maryland and was designated a naval aviator in May 2000 where he has accumulated nearly 2000 flight hours in various Navy aircraft.

Monagle's operational career includes two tours with Helicopter Mine Countermeasures Squadron Fourteen (HM-14) first as an airborne mine countermeasures (AMCM) Mission Commander, and later as a department head and the maintenance officer. As ship's company, Capt. Monagle served in the air department as the V-2 division officer and as a catapult and arresting gear officer ("Shooter") aboard USS Theodore Roosevelt (CVN-71). He also commanded the Blackhawks of Helicopter Mine Countermeasures Squadron Fifteen (HM-15) in Norfolk, Virginia.

His shore assignments include the MH-53E Fleet Replacement Squadron as a flight instructor and evaluator. Monagle has completed multiple acquisition tours with Naval Air Systems Command (NAVAIR) in Patuxent River, Maryland. First, as an assistant program manager for systems engineering (AMPSE), followed later as the H-60 Sustainment Integrated Product Team (IPT) Leader. He reported to the Defense Information Systems Agency's (DISA) in Fort Meade, Maryland to serve as military deputy PEO for Command and Control (C2). He led a team of Joint Service military members and DoD Civilians in modernizing C2 software systems for use by the joint services and coalition partners. He returned to NAVAIR to serve as the deputy program manager for Unmanned Carrier Aviation Mission Control System (UMCS).

Monagle is certified by the Project Management Institute as a project management and agile certified practitioner. He is a DoD-certified as an advanced practitioner in program management (Level III) and is designated as an acquisition professional.

Unmanned Low Profile Vessels (ULPVs): “Narco Subs” for Contested Logistics

Capt Sergio A. Sierra, USAF—Student, Naval Postgraduate School. [sergio.sierra@nps.edu]

Abstract

This research explores the potential military application of low-profile vessels (LPVs), also known as semi-submersible vessels (SSVs), commonly referred to as “narco subs,” which are extensively used by drug trafficking organizations (DTOs) for transporting illicit goods. LPVs’ effectiveness in evading interdiction is attributed to low observable attributes such as their aerodynamic shape, thermal shielding, and ability to ride very low in the water (minimal freeboard). LPVs can be manufactured affordably and quickly thanks to their simple design, easy-to-use building materials, and use of commercial-off-the-shelf (COTS) components. The research explores the concept of unmanned low-profile vessels (ULPVs) as a solution to contested logistics challenges within the U.S. military. This research aims to use modeling and simulation to analyze the idea of ULPVs supporting military logistics, offering insights into design considerations for an affordable, producible, and effective solution to enhance the U.S. military’s operational capabilities in a contested environment. In addition, this research intends to create an acquisition strategy for the DoD to leverage the U.S. industrial base, and potentially that of partner nations, to manufacture and field ULPVs affordably and at scale to meet DoD requirements. The final deliverables of this research effort are intended to provide the DoD with a consolidated product to inform decision making on questions regarding the military use of ULPVs.

Keywords: Narco, sub, low profile vessel, vessel, lpv, narco sub, unmanned, ulpv, unmanned low profile vessel, ssv, semi-submersible, ship, boat, design, digital modeling, simulation, contested logistics, logistics, cots, acquisitions, military, attritable, contested environment, usmc, usn, usaf, navy, marine corps, air force, eabo, dmo, ace, materiel, industry, small business, acquisition strategy

Introduction

The prevalent use of low-profile vessels (LPVs), also known as semi-submersible vessels (SSVs), and most referred to as “narco subs,” by drug trafficking organizations (DTOs), highlights a potentially advantageous model for a military capability, one largely untapped by the U.S. military. LPVs afford drug traffickers an affordable and effective means to move illicit material around the world, crossing vast distances of open ocean and evading some of the most sophisticated drug interdiction efforts aimed at preventing the successful transit of narco subs. The fundamental nature of LPVs to float minimally above the free surface has contributed to their effectiveness, while the simplicity of their construction and use of commercial off-the-shelf (COTS) technology has contributed to their affordability. Drug traffickers continually fabricate LPVs in the jungles and villages of South America to move their goods affordably and effectively throughout the world. In turn, the DTOs of South America have proven that LPVs are an effective and repeatable model, one that can be adapted to meet U.S. military requirements.

The use cases for LPVs in the U.S. military are likely many; however, an unmanned version of an LPV, an unmanned low-profile vessel (ULPV), may be an ideally suited materiel solution to address contested logistics challenges faced by the Joint force. The simplistic nature of LPV construction means that they can be constructed by a large portion of the U.S. and partner nation industrial bases, as opposed to only large defense industry shipyards. This large pool of potential manufacturers may result in manufacturing innovation and competition, further driving down material costs and introducing the ability to scale said production to high quantities compared to current U.S. shipbuilding capability. In this era of insufficient national shipbuilding capacity (Eckstein, 2024) and naval maintenance and repair backlogs (GAO, 2023), the ability to use alternative industrial sources is an essential requirement for any new approach.

Understanding how ULPVs can support contested logistics will benefit all Department of Defense (DoD) Services as each branch looks for options to maintain a sufficient logistics capability in a contested environment. In addition, understanding the technical and design considerations necessary for a ULPV to meet the requirements of the contested logistics mission, while also remaining affordable and simple enough to support high rates of production, will help inform the design of a desirable ULPV materiel solution to the DoD.

Problem

The problem is that, in the context of a hypothetical conflict with the People's Republic of China (PRC) in the Indo Pacific area of responsibility (AOR), there is a shortfall of logistics vessels to accomplish intra-theater logistics (Martin & Pernin, 2023). In addition, current logistics vessels are vulnerable against projected threats and are likely to be unescorted in a future large conflict (Larter, 2018), thereby negatively impacting projected success rates for vessels to deliver supply at their intended destinations. This capability gap is summarized as a lack of viable intra-theater logistics vessels, assuming an area denial, anti-access (A2AD) threat environment present from various weapon engagement zones (WEZs) from various PRC weapons systems deployed on land, air, and sea.

This matters because the foundation of any military to conduct operations hinges on successful logistics operations that provide the means to conduct military operations. Failure to address the capability gap in survivable intra-theater logistics vessels for military operations in the Indo Pacific will likely lead to a significant decrement in both the capacity and effectiveness of U.S. military operations, resulting in loss to U.S. persons, materiel, and objectives in the AOR.

This also matters because the existence of an intra-theater logistics capability gap undermines U.S. ability to deter military aggression or conflict escalation in the AOR. Rectifying the U.S. military's ability to confidently provide logistical support in a contested environment such as the Indo Pacific is a critical aspect of increasing its capacity for deterrence.

Purpose

The purpose of this research is to inform the design and employment of ULPVs to support military logistics operations in a contested environment like the Indo Pacific. This research also intends to create an acquisition strategy for the DoD to leverage the U.S. industrial base, and potentially that of partner nations, to manufacture and field ULPVs affordably and at scale to meet DoD requirements. The final deliverables of this research effort are intended to provide the DoD with a consolidated product to inform decision making on questions regarding the military use of ULPVs.

Scope

The geographic area of interest for this study begins at mainland China and extends to the expected maximum range of the DF-26B anti-ship ballistic missile WEZ, approximately 4,000km from the coast of mainland China, as depicted in Figure 1. This area contains the places of interest and the relevant distances therein for intra-theater logistics in the Indo Pacific.

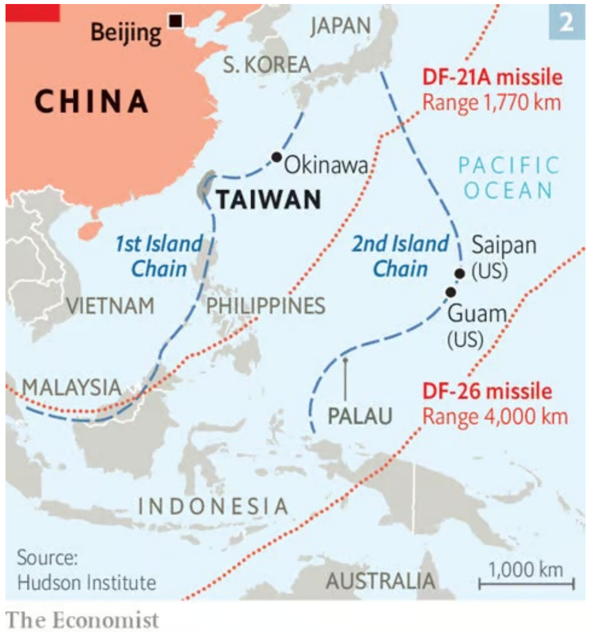


Figure 1. Indo Pacific Area of Interest & PRC Range Rings
("America and China Are Preparing for a War Over Taiwan," 2023)

Each of the services' operational models in the Indo Pacific are expected to be expeditionary in nature, thereby emphasizing forces that are mobile, agile, geographically distributed, and capable of various military operations within contested or potentially contested locations that may be austere or temporary in nature. The expected supply categories and their respective quantities anticipated for U.S. forces to conduct expeditionary operations in the Indo Pacific lay the foundation for the intra-theater logistical requirements. These logistical requirements inform the design of ULPVs intended to fill the AOR's intra-theater logistics capability gap. This study assumes that ULPVs may be designed to move any class of supply except for some specific class VII supply (major end items) that are anticipated to be too large and/or heavy for transport by ULPVs. The ability of a ULPV to carry any type and quantity of supply category is inherently limited by the design and function of the vessel, and as such, this research intends to consider the tradeoffs of notional ULPV designs and the resulting implications on the types and quantities of supply transportable.

This study assumes that ULPVs can complete these logistics functions for any unit of the U.S. military, regardless of service branch affiliation. This study also assumes that supply will need to be moved as break-bulk cargo and possibly include the use of shipping containers (and containers with similar form factors/MHE interfaces as shipping containers; i.e., tank containers) to move supply for military logistics functions.

Narco Subs: A Tool for Drug Trafficking

Drug trafficking organizations (DTOs) use various methods to traffic drugs by air, land, and sea. DTOs have historically innovated new means to traffic drugs as some prove more successful than others and as law enforcement agencies (LEAs) become more aware of and more effective at interdicting trafficking methods. One such innovative method used by DTOs is the use of "narco subs" to traffic drugs by sea. "Narco sub" is a term used to describe the three main categories of narco-vessels: Low Profile Vessels (LPV)/Self-Propelled Semi-Submersibles (SPSS), Submersibles/Fully Submersible Vessels (FSV), and Narco Torpedoes (the towed variety; Ramirez & Bunker, 2015). Most seized drug smuggling vessels to date are LPVs

(Ramirez & Bunker, 2015), and the focus of this research effort is on LPVs. LPVs cost DTOs approximately $1 million to manufacture and are built throughout Colombia and other parts of South America, in makeshift jungle boatyards (Figure 2), and in 30 to 45 days' time (VICE, 2011).

Figure 2. LPV Boatyard in Colombian Jungle
("The Archaeology of 'Narco Subs,'" 2020)

LPVs can carry up to 10 tons of drugs (Ramirez & Bunker, 2015) and can travel between 3,000 to 3,500 NM (VICE, 2011). In 2019, the first known trans-Atlantic crossing of a narco sub occurred when a 70ft LPV, carrying nearly 7,000lb of cocaine, made a 3,500-mile journey from Brazil to Spain (Figure 3) over a 27-day period (Jones, 2022). These vessels usually carry four crew members who make their voyage in very poor conditions, typically in a small aft space of the vessel that is hot, poorly ventilated, without a bathroom, and with makeshift bunking space (such as on top of fuel tanks; VICE, 2011).

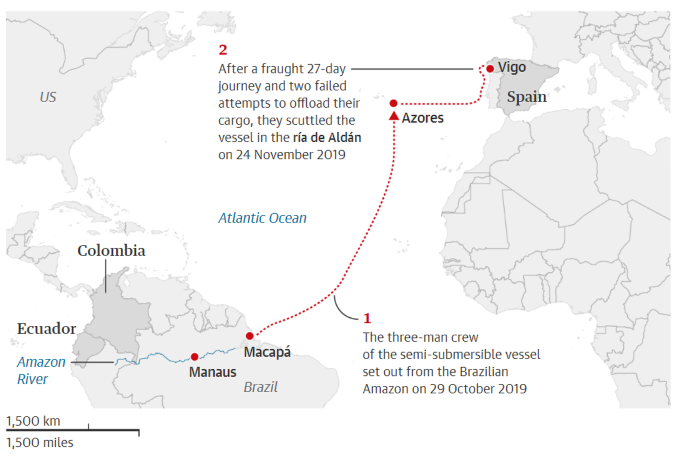


Figure 3. Trans-Atlantic Narco-Sub Journey
(Jones, 2022)

Generally, LPVs are difficult to detect as they are nearly impossible to spot from the horizon and very difficult to detect by radar (VICE, 2011). The low observable attribute of LPVs results from various design features such as the vessel: having minimal features on the deck, being aerodynamic in shape, riding very low in the water (minimal freeboard), using thermal shielding, being built of fiberglass, and being painted in a dark color that blends with the ocean surface (Figure 4; VICE, 2011).

Figure 4. View of Low-Profile Vessel Operating
(Sutton, 2021)

According to the Colombian Navy, however, one method of easily detecting LPVs, despite their lack of visible wake, is by spotting them from the air with an aircraft (VICE, 2011). A 2014 account by U.S. Navy CAPT Mark F. Morris supported the need for aircraft utilization to achieve favorable LPV detection probability, stating:

> American operations analysis shows that given good intelligence of a drug event and a patrol box of a certain length and width, a surface vessel operating alone has only a 5% probability of detecting (PD) that event. A surface vessel with an embarked helicopter increases the PD to 30%, and by adding a Maritime Patrol Aircraft to the mix, the PD goes up to 70%. Analysis by the Colombian Navy shows that adding one of their submarines to the mix raises the PD to 90%. (Ramirez & Bunker, 2015)

The U.S. Drug Enforcement Administration has considered that only about 20% of narco subs are intercepted (Ramirez & Bunker, 2015). In a 2014 testimony to Congress, U.S. Southern Command (SOUTHCOM) reported that low interdiction rates were due to asset shortfalls (Ramirez & Bunker, 2015), presumably resulting in an inadequate number of vessels and aircraft able to conduct maritime interdiction missions against LPVs. Most narco subs have been found in the SOUTHCOM area of responsibility (AOR), with 78% being found in the Pacific (in waters near South and Central America) and 20% being found in the Caribbean (Ramirez & Bunker, 2015). As a result, most LPV interdiction data exists in an environment where LEAs are under-resourced, according to the 2014 SOUTHCOM testimony to Congress, resulting in uncertainty at how effective LPVs are at avoiding detection and interdiction in an environment where they are hunted with more numerous resources.

For DTOs, the business model of LPV fabrication and operation is the result of a cost-benefit analysis where the yielded benefits are far superior to the costs associated with building and operating LPVs (Ramirez & Bunker, 2015). A 10-ton cargo of narcotics may be worth approximately $200 million (Ramirez & Bunker, 2015), minus the $1 million construction cost of the LPV, which leaves a $199 million profit per successful LPV voyage. Factoring in a loss rate of 20%, based on the previously mentioned LPV interdiction rate, results in an average profit per

LPV voyage of approximately $159 million. This calculation assumes a full 10-ton cargo on every LPV voyage as well as a constant interdiction rate of 20%; however, it serves to highlight the superior benefit over the cost of LPV fabrication and operation, resulting in the continued DTO use of LPVs for drug trafficking.

The Appeal of LPVs for Contested Logistics

In a foreword to *Beans, Bullets, and Black Oil*, former Secretary of the Navy Dan A. Kimball highlighted the criticality of logistics to the fight against the Japanese Empire in World War II, saying:

> Victory is won or lost in battle, but all military history shows that adequate logistic support is essential to the winning of battles. In World War II, logistic support of the fleet in the Pacific became a problem of such magnitude and diversity, as well as vital necessity, that all operations against Japan hinged upon it. (Carter, 1998)

Given the success that DTOs experience trafficking drugs with LPVs, it is fair to question if a vessel like an LPV could be used in a military logistics role for the DoD. This question exists at a time when the United States prepares for a possible conflict in the Indo Pacific between China and Taiwan, at a time when the commander of the U.S. Pacific Fleet warned of an insufficient Combat Logistics Force (Katz, 2024). Wargames indicate that U.S. logistics vessels will be sought after by any adversary (Katz, 2024), and past exchanges with Chinese naval leadership indicate that these vessels will be primary targets in a U.S.–China conflict (Suciu, 2020). Joint U.S. forces will require sustainment to effectively fight a war in the Indo Pacific, and that sustainment must ensure support that flows from the United States to the point where U.S. Transportation Command delivers supplies and further to the point where frontline forces receive supplies (Martin & Pernin, 2023). The logistics supply chain in this case spans the geographic distances between factories within the continental United States to military forces staged throughout the Indo Pacific. Martin and Pernin (2022) highlight that the most particularly concerning stretch of the logistics map from the United States to the frontlines of the Indo Pacific is the part known as intra-theater lift, "the portion of the transportation chain that delivers materiel from a port of debarkation to the point of use by an operational unit."

Although individual services have capabilities to meet a portion of their intra-theater transportation demands, when combined, they do not meet all needs of the joint force (Martin & Pernin, 2023). In addition to the sheer quantity of supply that would need to be transported across large distances over water, a fight in the Indo Pacific would leave U.S. logistics vessels to contend with growing anti-access/area denial (A2/AD) capabilities of the PRC. These PRC capabilities span air, land, and sea, and leverage various missiles of growing quantity and capability intended to impose maximum attrition to slow and impede any adversarial military operations (Joshi, 2019). PRC A2/AD capabilities would envelop the entirety of what will be the intra-theater logistics operating area for a U.S. military operation in the Indo Pacific (Joshi, 2019). Because logistics operations are expected to take place in contested environments, and because the DoD lacks the logistics forces to support a large military campaign in the Indo Pacific, the need for new materiel solutions to accomplish contested logistics missions has arisen (Mills & Limpaecher, 2020).

One thought to help address the capability gap in intra-theater contested logistics is to apply the DTO model of LPVs to U.S. military logistics, perhaps even in an unmanned capacity (Mills & Limpaecher, 2020). Narco-sub–like vessels such as LPVs are thought of as a prospective materiel solution to provide logistics support to the USMC's expeditionary advanced base operations (EABO; Mills et al., 2020) or to Taiwan in the event of a Taiwan conflict (Griffin, 2024). A U.S. unmanned low-profile vessel (ULPV) may be able to leverage the low observable benefits that make DTO-operated LPVs difficult to detect and interdict but without the need of a crew and subjecting that crew to the conditions and risks associated with an LPV operating in the waters of the Indo Pacific under the PRC's A2/AD threat bubble.

DTO LPV Design Themes

There are a few key design themes that arise from DTO LPVs that are foundational to the success of the drug trafficking LPV model. These design themes are design simplicity, design for mission needs, and design for asset attritability. Narco subs evolved over decades, beginning in the early 1990s with experimentation, through the early 2000s with prototyping, and continuing from 2007 to the present with design standardization and maturation (Ramirez & Bunker, 2015). One similarity among photos of all captured or interdicted narco subs is the simplicity of design that they all share (Ramirez & Bunker, 2015). Shaping wood and fiberglass into a functional LPV within 30 to 45 days, using local unskilled labor (VICE, 2011), in the jungles of Colombia is possible because of simple vessel design. Perhaps assisting the rapid LPV manufacture timeline is what Ramirez & Bunker (2015) indicate, that DTOs use readily available commercial off-the-shelf (COTS) components for the engines, navigation systems, and communications systems for their LPVs. The interiors of these LPVs further highlight their design simplicity, with little to no accommodations made for the crew and the sole focus on mission needs like cargo carrying capacity (large cargo holds) and vessel range (large fuel tanks), with neither compromised to carve out space for the crew (Figure 5). In a sense, DTOs have created a minimal viable product (MVP) to accomplish maritime drug trafficking at the lowest possible cost and highest possible benefit.

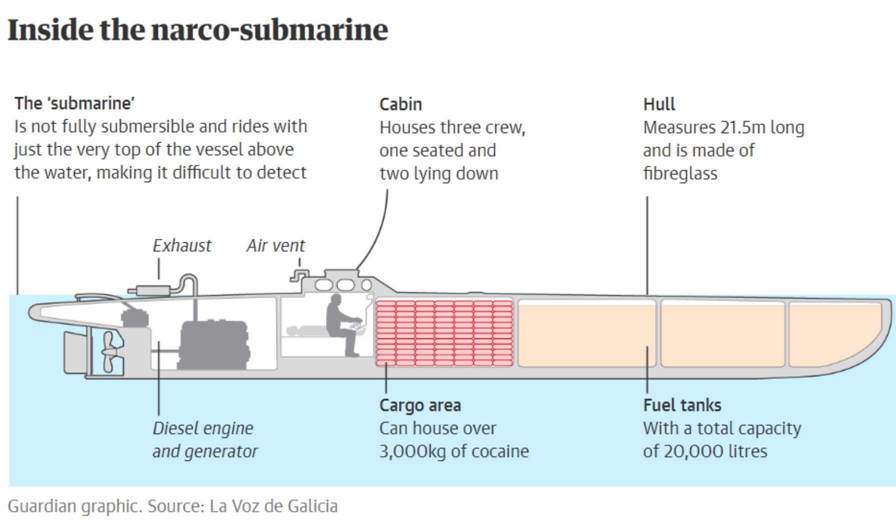


Figure 5. Cutaway of LPV Highlighting Crew, Cargo, Engine, and Fuel Spaces
(Jones, 2022)

In addition to design simplicity, LPVs appear tailor-designed for their mission needs. As LPVs have evolved over time, their design has become more aerodynamic, they have less piping on the hull, they run awash or with less freeboard, they incorporate lead shielding, and they use seawater to cool exhaust gases, all to decrease the probability of detection by counter-drug operation by LEAs (VICE, 2011). While the hulls have become more aerodynamic and larger in size, their shapes continue to remain like a sealed 'go-fast' boat with a deep V-shaped hull, sufficient for the sea states they operate in (Ramirez & Bunker, 2015). Within the confines of this hull design, maximum space is afforded for cargo and fuel capacity. Loading and unloading the LPVs is accomplished through a simple single hatch on the vessel, by hand, and either at dock or at sea (VICE, 2011).

One additional theme to highlight is the inherent attritable nature of LPVs manufactured and operated by DTOs. These LPVs include a scuttle valve that floods the hull if activated by the crew (VICE, 2011), and it is used often in LPV interdictions to prevent LEAs from obtaining criminal evidence (Ramirez & Bunker, 2015). Even if the LPV reaches its destination and successfully unloads its cargo, LPVs are typically scuttled rather than reused (VICE, 2011). Since LPVs are typically only valued at 2–3% the value of the cargo they carry, they are viewed as expendable (VICE, 2011).

It is also worth noting that DTO LPV designs are unbounded by regulations on maritime transport, such as those governed by the International Maritime Organization (IMO; n.d.). The IMO sets standards for the safety, security, and environmental performance of international shipping (IMO, n.d.), and it is likely that the acquisition process for any sort of LPV or ULPV by the DoD would need to comply with maritime specifications, standards, and laws for vessel design, construction, and operation, all factors that are not concerning to DTOs.

Considerations for ULPVs in the DoD

The question at hand is how the DoD could best adopt the DTO LPV model and evolve it to an unmanned asset that can accomplish logistics operations for the joint force in a contested environment. Helpful to addressing this question is framing it within the context of an operational area where ULPVs may be utilized, such as the Indo Pacific in a notional conflict with the PRC. The following subject areas are an overview of several key considerations that the DoD should consider for the design and operation of ULPVs in the Indo Pacific.

Initial Thoughts: Vessel Design

LPVs are immersed more than standard surface vessels; however, maintaining a minimal freeboard and proximity to the free-surface allows LPVs to use low-cost combustion engines while also negating the need for costly pressure vessels, submarine control surfaces, and other mechanisms necessary for a vessel that operates fully submerged (Sung et al., 2022). Initial analysis indicates that LPVs have increased stability with more slender hull shapes (Sung et al., 2022), and a review of DTO LPVs shows a trend toward increasingly slender vessels over time (Sutton, 2020).

It is important to consider the differences in sea conditions, or sea states, that exist between the waters where DTOs operate versus the waters of the Indo Pacific. As there is little to no data on how LPVs would perform in the sea states of the Indo Pacific, some initial research has been done on semi-submersible vessels (SSVs) which can be applied to LPVs. Initial research at the U.S. Naval Academy indicates that LPV hydrodynamic performance would be very sensitive to the forces of surface waves and that more extensive testing is needed (Sung et al., 2023). Further findings include increased resistance, due to increased hydrodynamic drag, experienced with a hull operating more immersed (Sung et al., 2023), likely equating to a need for greater power requirements than traditional surface vessels to attain a similar operating speed.

In addition to hydrodynamic considerations, most of which require further research for DoD adoption of SSVs (and LPVs; Sung et al., 2023), LPV design should also consider the material choice for fabrication as well as the complexity of the vessel design. As previously discussed, drug trafficking LPVs are typically made of wood and fiberglass. These materials are more affordable and easier to build with compared to metal, requiring less skilled labor or specialized machinery. In addition, these materials are harder to detect with radar than metal is. In the context of military conflict, these materials may be advantageous to help defeat threats that ULPVs encounter in the waters of the Indo-Pacific. Maintaining a vessel design that is simple and with as few extra features or building steps as possible will allow the DoD to follow the DTO LPV model of minimal cost, thereby driving towards a design that is both affordable and able to be rapidly built, increasing the chance of the ULPV being considered attritable.

Finally, it should be kept in mind how the vessel is intended to be loaded, unloaded, and interfaced with by people and other vessels or equipment. For a vessel with the primary mission of transporting supply for logistics, it is paramount that the vessel be designed with the operational environment in mind. For example, if the ULPV needs to resupply Marines operating on expeditionary advanced bases (EABs) in the Indo Pacific and the island EAB location does not have a pier, then it should be considered if the ULPV needs to be able to beach or if the Marines will have to retrieve the supply by other means. If the ULPV needs to be able to beach, it must be able to make its way through the shallow, and often reef- and rock-strewn, water of islands in the Indo Pacific. This requires a vessel with a shallow draft, a hull attribute poorly suited for transiting rough sea states over large distances. DTO LPVs do not have a shallow draft hull; then again, DTO LPVs typically load and unload at sea or pier side. It should also be considered how supply will be loaded onto and off the ULPV, either by crane, roll-on/roll-off, by hand, or otherwise.

Other Considerations Being Researched

This research effort is exploring various other considerations for ULPVs in the DoD in addition to the few mentioned above. The final report on this effort will include greater detail on all of the aforementioned areas. Other considerations that will be included in the final report include but are not limited to vessel performance, autonomy, supply types, vessel loading, vessel unloading, vessel interdiction and tampering, external communications, positioning, navigation, and timing (PNT), and command and control (C2).

Possible ULPV Designs

This research is aware of multiple designs for ULPVs that will be analyzed. One such design, from CDR Todd Greene at the U.S. Naval Academy, is called the NightTrain (Greene, 2023). NightTrain is an innovative ULPV concept with a unique design to ferry shipping containers across large distances, proposing to move supply from the factory to the frontlines (Greene, 2023). Other ULPV designs resemble DTO LPVs, only without the need for a crew. Estimated performance parameters of these designs will be utilized in this research's modeling and simulation efforts to provide expected performance parameters, such as a design's probability of detection.

Modeling & Simulation

Part of this research aims to use modeling and simulation to help inform relatively unknown aspects concerning the idea of ULPVs for military contested logistics. One aspect of analysis seeks to understand the probability of detection and susceptibility of ULPVs operating in the Indo Pacific against threats from the PRC. Another aspect of analysis seeks to understand the impact of ULPVs on maintaining a steady level of supply for expeditionary units, such as Marines operating on EABs during a conflict with the PRC. Yet another aspect of analysis will incorporate virtual sandboxing via a virtual sand table (VST) to visually depict ULPV employment and collect data to inform potentially new considerations for ULPVs.

Next Generation Threat System (NGTS)

"NGTS is a military simulation environment produced by the Naval Air Warfare Center Aircraft Division (NAWCAD) that provides real-time military scenario simulations. NGTS models threat and friendly aircraft, ground, surface, subsurface platforms, corresponding weapons and subsystems, and interactions in a theater environment" (Tryhorn et al., 2023). NGTS modeling and simulation work in this research is a collaborative effort between the Naval Postgraduate School and the Naval Information Warfare Center Pacific (NIWC Pacific).

NGTS will be used to simulate the performance of ULPVs in a contested environment, specifically, the ability of ULPVs to transit a body of water shared by various types and quantities of PRC surface and airborne assets without being detected or destroyed. NGTS will be used to simulate two types of environments in the contested space making up intra-theater lift, open water ("blue" water) transit and littoral transit, as the type and quantity of PRC assets encountered in each environment are likely to be different. Many cycles of NGTS simulations will run to output data on a ULPV's probability of detection in an Indo Pacific conflict. As ULPVs are assumed to be unarmed in this research, their greatest chance of successfully completing logistics missions in a contested battlespace may rely on remaining undetected.

NGTS Assumptions & Limitations

Some assumptions and limitations are made in the NGTS simulation. First, the data used to create the performance parameters for all red team (PRC) assets is based on unclassified, open-source information. Second, the performance parameters of blue force (U.S.) logistics vessels are either based on open-source information or estimated based on similar existing vessel parameters (such as parameters of DTO LPVs to inform some ULPV parameters). Third, the type and kind of red force assets present in each of the operational environments (blue water and littoral) are best estimates based on open-source information of PLA order of battle data. Other NGTS assumptions and limitations will be documented as this research's modeling and simulation effort progresses.

Causal Loop Diagram

The NGTS modeling and simulation work will inform ULPV probability of detection given specific red force (PRC) capabilities based on various asset types and quantities. The probability of detection data output from the NGTS simulation runs will then be input into a causal loop diagram (CLD) to simulate the larger interaction of variables concerning ULPVs maintaining a level of supply at an expeditionary base location. "Causal loops diagrams (also known as system thinking diagrams) are used to display the behavior of cause and effect from a system's standpoint. A CLD is a causal diagram that aids in visualizing how different variables in a system are interrelated" (Barbrook-Johnson & Penn, 2022). According to Barbrook-Johnson and Penn (2022),

> CLD are made up of connections, or edges, which represent causal influence from one node to the other; either positive (i.e. they increase or decrease together) or negative (i.e. they change in opposite directions, if one goes up, the other goes down, and vice versa). The maps always show and focus on feedback loops, both in the construction of the map and in its visualization. Loops are made conspicuous by the use of curved arrows to create circles.

This research effort will use the following CLD (Figure 6) or a version of it (as this research is still ongoing and the following CLD is still a work in-progress) to analyze the impact of ULPVs on maintaining a steady level of supply for expeditionary units. The following CLD is a draft, working product.

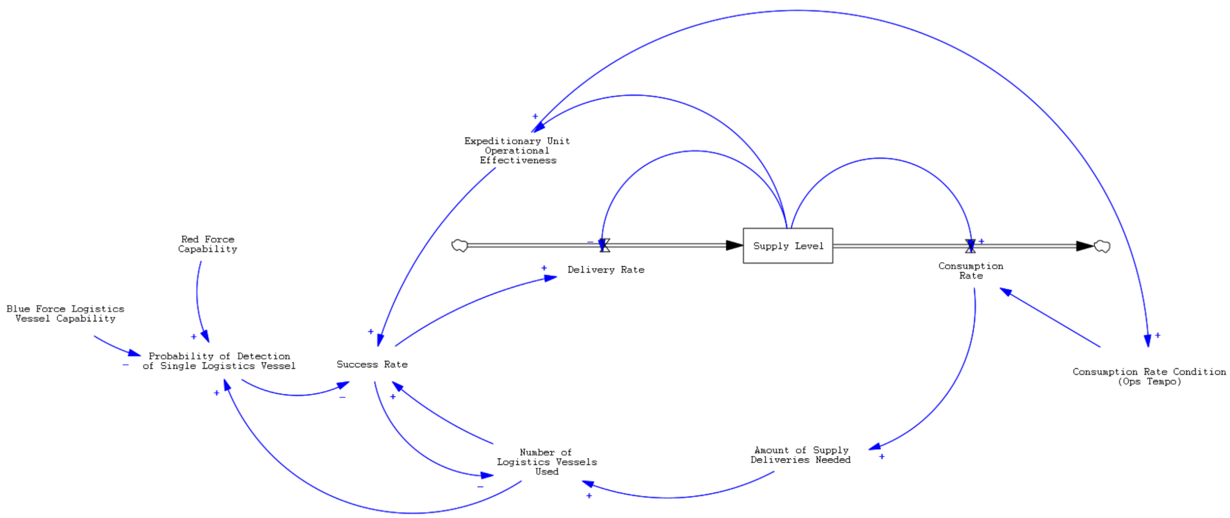


Figure 6. Causal Loop Diagram for Contested Logistics and Expeditionary Unit Resupply

Model Variable Definitions Listed below are definitions of the variables being utilized in this research effort's modeling and simulation:

- Blue Force Logistics Vessel Capability: The collective attributes of the blue force (U.S.) logistics vessel in question contributing to its detectable signature.
- Red Force Capability: The collective attributes of the red force (PRC) assets (surface vessels and airborne craft) resulting from the type and quantity of red force assets attempting to detect and destroy blue force logistics vessels.
- Probability of Detection of Single Logistics Vessel: Probability that the logistics vessel in question will be detected by red force assets.
- Success Rate: Probability that the logistics vessel in question will not be detected, interdicted, or destroyed by the red force and will therefore reach its delivery destination.
- Delivery Rate: The number of deliveries per measure of time.
- Expeditionary Unit Operational Effectiveness: The ability of an expeditionary unit to support its own needs to maintain unit health, readiness, and the ability to successfully complete any tasked mission.
- Supply Level: The amount of various supply classes that must be maintained at an operational unit to support that unit's health, readiness, and ability to successfully complete any tasked mission.
- Consumption Rate: The amount of supply consumed per measure of time.
- Consumption Rate Condition (Ops Tempo): The influence exerted on the consumption rate given the level of operational activity intensity at a point in time.
- Amount of Supply Deliveries Needed: Quantity of resupply missions required (based on supply capacity of logistics vessel in question) to replenish supply level at expeditionary unit.
- Number of Logistics Vessels Used: Quantity of logistics vessels utilized to resupply expeditionary unit.

CLD Assumptions & Limitations Some assumptions and limitations are made in this CLD. First, the logistics vessels in question will be unarmed and defenseless. Second, logistics vessels will carry supplies that are of equal type and proportional quantity necessary to maintain the total supply level stock at the expeditionary unit. Other CLD assumptions and limitations will be documented as this research's modeling and simulation effort progresses.

ULPV Specific CLD: ULPV System Performance & Expeditionary Unit Resupply The previous CLD illustrates the conceptual interactions between a logistics vessel, the red forces, and the expeditionary unit supply level. However, in the case of a fully unmanned system such as a ULPV, additional factors would play into the CLD to illustrate the need for the ULPV as a series of systems to perform as desired. The resulting desired system performance would then be a variable, in addition to the probability of detection, impacting the success rate of the vessel. In addition, a presumed reliance on external connectivity between the ULPV and any logistics command and control structure would introduce another series of variables that contribute to the desired system performance as well as the vessel's RF signature.

This research effort may/may not pursue simulation of a CLD specific to a ULPV. However, it is useful to see the possible interactions between CLD variables specific to a ULPV. Below (Figure 7) is a draft, working version of a possible ULPV CLD.

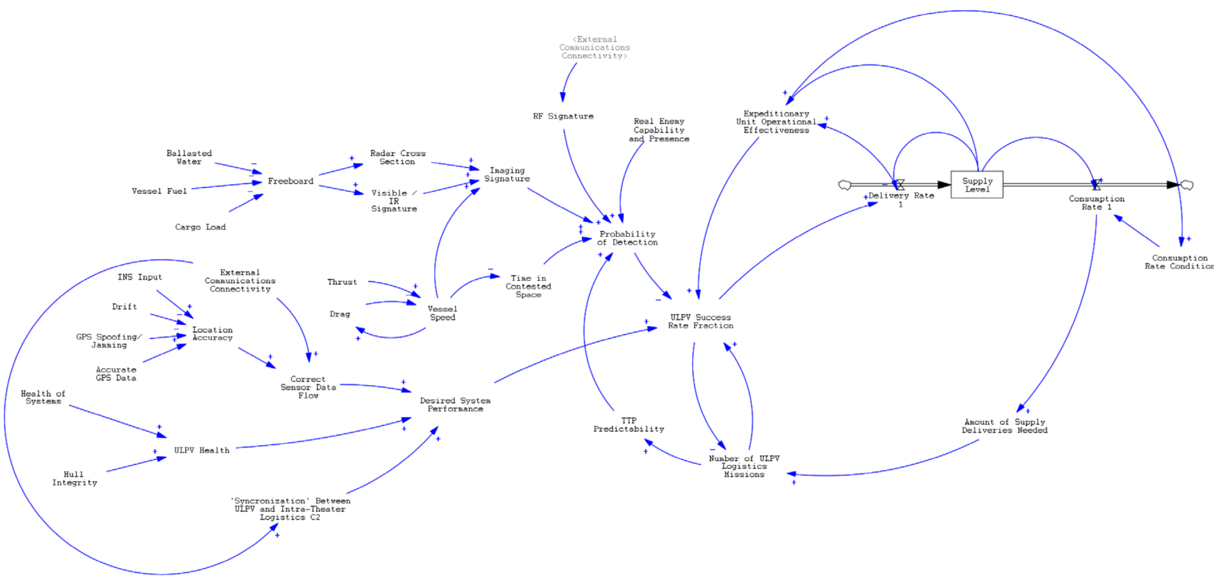


Figure 7. ULPV Specific Causal Loop Diagram for Contested Logistics and Expeditionary Unit Resupply

Virtual Sandboxing

The Modeling Virtual Environments and Simulation (MOVES) Institute at the Naval Postgraduate School has a virtual sand table (VST; Figure 8) that will be utilized with 3D printed models of notional ULPVs to visually depict ULPV employment on real-world projected locations. This process will enable data collection to inform potentially new considerations for ULPV design and employment. 3D models used in SPIDERS3D are produced from a variety of sources, then converted (if necessary) to X3D for mashup composition and Web-based collaborative visualization. Extensible 3D (X3D) Graphics is the international standard for publishing interactive 3D models on the Web (Brutzman & Daly, 2007). More information is available at https://www.web3d.org/x3d/what-x3d.

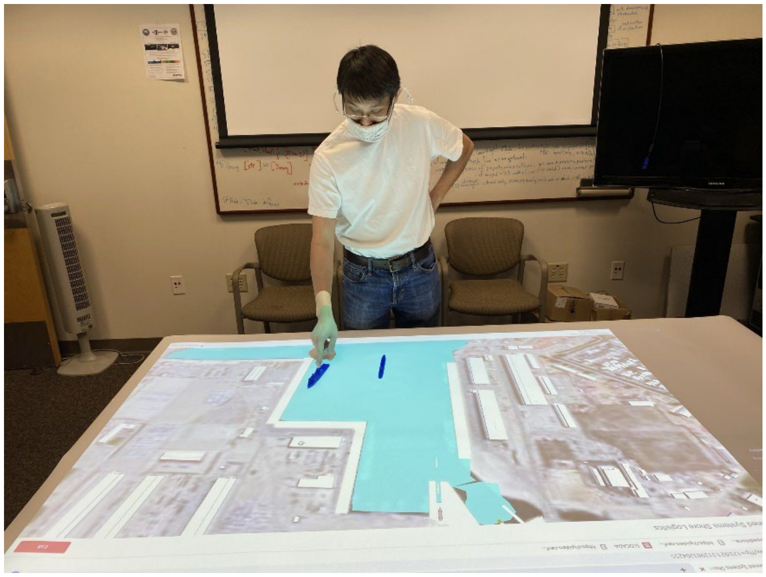


Figure 8. Demonstration of Virtual Sand Table at NPS
(EXWC SPIDERS3D, 2021)

ULPV Acquisition

In addition to the documenting design and employment considerations of ULPVs to support military logistics operations in a contested environment, this research also intends to create an acquisition strategy for the DoD to leverage the U.S. industrial base, and potentially that of partner nations, to manufacture and field ULPVs affordably and at scale to meet DoD requirements. This acquisition strategy will consider all LPV design considerations and lessons learned from DTO LPV operations as well as findings from this research's modeling and simulation efforts.

Considerations for ULPV Production

This research also intends to analyze various considerations to producing ULPVs, especially those expected to impact vessel cost and vessel production time, as part of an overarching intent to analyze the ULPV's ability to be used as a surgable, sustainable, and attritable materiel solution in support of national defense responsibilities to deter, de-escalate, and defeat. One interesting consideration for ULPV production is the prospect of leveraging small businesses and boatyards throughout the United States, vice shipyards, given the insufficient national shipyard capacity (Eckstein, 2024) that may be unlikely to meet production demands of a new line of vessels such as ULPVs. There may exist a relationship between ship design simplicity, COTS component utilization, and material choices that result in a level of production complexity not outside the capability of many small businesses and boatyards throughout the United States. Further, ULPV designs and their respective production complexities may or may not easily support production of ULPVs in host or partner nations throughout the Indo Pacific. The ability to produce ULPVs within the theater of conflict would save the use of copious resources needed to transport these vessels into theater. The analysis of this research intends to inform the ease with which ULPVs may be produced in the Indo Pacific.

Closing Thoughts

ULPVs may provide the DoD long-term stabilization value during an era of grey-zone competition and military conflict. Having a contested logistics capacity to provide indefinite logistics resupply across first and second island chains in the Indo Pacific provides paths for de-escalation back to deterrence, rather than unchecked escalation to conflict. The Liberty Ship program of World War II proved critical to the war's outcome (Herman, 2012). Liberty Ships overcame attrition by German U-Boats in the contested waters of the Atlantic. Many lessons learned from the design and production of Liberty Ships can similarly inform the design and production of ULPVs to overcome threats in the contested waters of the Indo Pacific. Some applicable lessons learned include utilizing principles of standardization and methods of mass production (Lane, 1951, pp. 31, 72) as well as the use of machine tools and prefabrication (Herman, 2012). These same lessons were applied with great success by Andrew Higgins in 1942 (Lane, 1951, p. 185), resulting in the design and mass production of "Higgins boats," tens of thousands of which were shallow-draft landing craft made of wood and steel for amphibious assaults in the Indo Pacific (Strahan, 1994). The geography of the Indo Pacific since World War II remains very similar today, and applying lessons learned from vessels designed, produced, and employed during World War II may prove beneficial to inform ULPV design, production, and employment for the DoD today.

Acknowledgements

This research would not be possible without the guidance and support of many individuals. The scope of this research is multi-disciplinary, and it is the product of collaboration between various people who are passionate about the defense of the United States and its interests.

This research benefits greatly as a sponsorship effort by the Naval Information Warfare Center Pacific (NIWC Pacific) as part of the NIWC Pacific fellowship between the Naval Postgraduate School (NPS) and NIWC Pacific. Thank you to Michael Stuckenschneider for advising this research, being a valuable resource, and connecting me to the right people. Additionally, thank you to members of the NIWC Pacific team assisting this research's modeling and simulation efforts, especially Dr. Glenn-Iain Steinback.

This research is part of the NPS' Nimitz Research Group (NRG) and benefits from that association. The NRG trip to PACFLT in February 2023 sparked the conversations that directly led to this research effort. Thank you to Vice Admiral Phil Sawyer, USN (Ret.) for supporting this research.

Thank you to all the advisors supporting this dual-degree masters' thesis: Dr. Don Brutzman, Dr. Jarema Didoszak, Professor Ray Jones, and Professor Howard Pace. Thank you for supporting the many meetings regarding this research. In addition, I would like to extend my thanks to the following NPS faculty for supporting this research: Dr. Mustafa Canan, Dr. Sean Kragelund, Professor Christian Fitzpatrick, and Professor Jeffrey Kline.

Finally, thank you to Dr. Diogenes (Dio) Placencia for always being very helpful and responsive and connecting me to the right people to address various questions.

References

America and China are preparing for a war over Taiwan. (2023, March 9). *The Economist*. https://www.economist.com/briefing/2023/03/09/america-and-china-are-preparing-for-a-war-over-taiwan

The archaeology of "narco subs." (2020, October 2). *JaySea Archaeology*. https://jayseaarchaeology.wordpress.com/2020/10/02/the-archaeology-of-narco-subs/

Barbrook-Johnson, P., & Penn, A. S. (2022). Causal loop diagrams. In P. Barbrook-Johnson & A. S. Penn, *Systems Mapping* (pp. 47–59). Springer International Publishing. https://doi.org/10.1007/978-3-031-01919-7_4

Brutzman, D., & Daly, L. (2007, April). *X3D: Extensible 3D graphics for web authors*. Elsevier. https://dl.acm.org/doi/book/10.5555/1214715

Carter, W. R. (1998). *Beans, bullets, and black oil: The story of fleet logistics afloat in the Pacific during World War II*. Naval War College Press.

Eckstein, M. (2024, March 20). *Navy 30-year shipbuilding plan relies on more money, industry capacity*. Defense News. https://www.defensenews.com/naval/2024/03/20/navy-30-year-shipbuilding-plan-relies-on-more-money-industry-capacity/

EXWC SPIDERS3D. (2021, March 14). *Virtual sand table.* Naval Postgraduate School. GitLab. https://gitlab.nps.edu/Savage/Spiders3dPublic/-/blob/master/videos/demonstrations/VirtualSandTable/README.md

GAO. (2023, May 4). *Navy ships: Applying leading practices and transparent reporting could help reduce risks posed by nearly $1.8 billion maintenance backlog.* https://www.gao.gov/products/gao-22-105032

Greene, T. (2023, June 8). *The NightTrain: Unmanned expeditionary logistics for sustaining Pacific operations.* Center for International Maritime Security. https://cimsec.org/the-nighttrain-unmanned-expeditionary-logistics-for-sustaining-pacific-operations/

Griffin, P. (2024, January 25). *Contested logistics: Adapting cartel submarines to support Taiwan*. U.S. Naval Institute. https://www.usni.org/magazines/proceedings/2024/january/contested-logistics-adapting-cartel-submarines-support-taiwan

Herman, A. (2012, May 8). *Freedom's forge: How American business produced victory in World War II* (2013 Random House trade paperback edition). Random House Trade Paperbacks.

International Maritime Organization. (n.d.). *Introduction to IMO*. https://www.imo.org/en/About/Pages/Default.aspx

Jones, S. (2022, February 4). In too deep: The epic, doomed journey of Europe's first narco-submarine. *The Guardian*. https://www.theguardian.com/society/2022/feb/04/europe-first-narco-submarine-agustin-alvarez

Joshi, S. (2019, April 10). Demystifying the anti-access/area denial (A2/AD) threat. *Medium*. https://sameerjoshi73.medium.com/demystifying-the-anti-access-area-denial-a2-ad-threat-d0ed26ae8b9e

Katz, J. (2024, February 20). Navy's Combat Logistics Force on "narrow margins," US Pacific Fleet chief warns. *Breaking Defense*. https://breakingdefense.sites.breakingmedia.com/2024/02/navys-combat-logistics-force-on-narrow-margins-us-pacific-fleet-chief-warns/

Lane, F. C. (1951). *Ships for victory; A history of shipbuilding under the United States Maritime Commission in World War II*. Johns Hopkins Press.

Larter, D. (2018, October 11). *'You're on your own': US sealift can't count on Navy escorts in the next big war*. Defense News. https://www.defensenews.com/naval/2018/10/10/youre-on-your-own-us-sealift-cant-count-on-us-navy-escorts-in-the-next-big-war-forcing-changes/

Martin, B., & Pernin, C. G. (2022). *The problem of intra-theater lift: Moving things around in the Pacific area of responsibility*. https://www.rand.org/pubs/commentary/2022/09/the-problem-of-intra-theater-lift-moving-things-around.html

Martin, B., & Pernin, C. G. (2023, June 23). *So many questions, so little time for Pacific logistics*. https://www.rand.org/blog/2023/06/so-many-questions-so-little-time-for-pacific-logistics.html

Mills, W. D., & Limpaecher, E. (2020). *Sustainment will be contested*. https://calhoun.nps.edu/handle/10945/66167

Mills, W., Phillips-Levine, D., & Fox, C. (2020, July 22). *"Cocaine logistics" for the Marine Corps*. War on the Rocks. https://warontherocks.com/2020/07/cocaine-logistics-for-the-marine-corps/

Ramirez, B., & Bunker, R. J. (2015, February 2). *Narco-submarines: Specially fabricated vessels used for drug smuggling purposes.* Foreign Military Studies Office. 1–164.

Strahan, J. E. (1994). *Andrew Jackson Higgins and the boats that won World War II*. Louisiana State University Press.

Suciu, P. (2020, June 9). *The really boring way China would try to win a war against America*. The National Interest, The Center for the National Interest. https://nationalinterest.org/blog/buzz/really-boring-way-china-would-try-win-war-against-america-162036

Sung, L. P., Laun, A., Leavy, A. J., Ostrowski, M., & Postma, M. (2022). Preliminary hull-form design for a semi-submersible vessel using a physics-based digital model. *Naval Engineers Journal*, *134*(4), 65–71.

Sung, L. P., Matveev, K. I., & Morabito, M. G. (2023). Exploratory study of design parameters and resistance predictions for semi-submersible vessels. *Naval Engineers Journal*, *135*(1), 185–203.

Sutton, H. (2020). *Narco submarines: Covert shores recognition guide*. San Bernardino. https://community.apan.org/wg/tradoc-g2/fmso/m/fmso-monographs/197161

Sutton, H. I. (2021, September 1). *Colombian Navy interdicts super low-profile narco submarine*. Covert Shores. http://www.hisutton.com/New-Type-Of-Narco-Submarine-2021-09-01.html

Tryhorn, D., Dill, R., Hodson, D. D., Grimaila, M. R., & Myers, C. W. (2023). Modeling fog of war effects in AFSIM. *The Journal of Defense Modeling and Simulation*, *20*(2), 131–146. https://doi.org/10.1177/15485129211041963

VICE. (2011, October 26). *Mother board | Colombian narcosub*s [Video]. YouTube. https://www.youtube.com/watch?v=2Rp-C1ph_g8&t=94s&ab_channel=VICE

System Product Line Cost and Investment Modeling Applied to UUVs

Raymond J. Madachy—is a Professor with the Systems Engineering Department at the Naval Postgraduate School. [rjmadach@nps.edu]

John M. Green—is a senior lecturer with the Systems Engineering Department at the Naval Postgraduate School. [jmgreen@nps.edu]

Abstract

Integrated cost and product modeling applied to the acquisition of unmanned underwater vehicles (UUVs) demonstrated the economic benefits of a product line strategy. The modeling framework includes system modeling language (SysML) for product modeling and a constructive cost model set. The constructive product line investment model (COPLIMO) framework was used for return on investment (ROI) analysis with the constructive systems engineering cost model (COSYSMO) for single system investment and reuse costs. Cost model inputs were extracted directly from the SysML requirements and executable activity models for the UUVs. Model integration reduces effort since only product modeling is performed without the need for independent cost modeling expertise.

The case study research investigated the reduction of acquisition costs applying the integrated product line acquisition model for UUV missions with overlapping requirements. The key research question focused on the ROI of a product line approach for UUV systems developing a baseline architecture for reuse. Supporting questions addressed the reuse savings for individual UUV systems, the size and complexity of the resulting system, and their estimated effort. Results indicate a strong ROI when using a product line approach for UUV systems.

Keywords: product lines, economics, COPLIMO, COSYSMO, cost modeling, ROI, UUV, systems engineering

Introduction

The product line engineering concept (PLE) integrates well with the adaptive acquisition framework introduced in the fall of 2020. Because the PLE is based on the concept of a common platform that can be used to develop a family of products It offers the capability to reduce acquisition cost and “time to market.” PLE is based on a two–life cycle model That integrates the domain of interest with relevant applications. This facilitates the development of systems through the identification of commonalities and system variabilities. This premise is the basis for the application of the constructive product line investment model (COPLIMO) framework to case studies with the intent of developing a viable cost modeling methodology that would support the adaptive acquisition framework.

Active student research (group capstones and individual theses) on combat system product line architectures and costs using model-based systems engineering (MBSE) methods with COPLIMO variants have been applied and extended across Naval domains at NPS (Table 1).

Table 1. Naval Case Studies

System Case Study	Sizing Unit(s)	Equivalent Size Adjustments	Reuse and Investment Model	MBSE Models	Empirical Data Used	Baseline System Size for Analysis
Cruise Missile Tiers	system component	reuse category	Basic COPLIMO	OVM, data flows	subsystem costs	20 subsystems
Aegis Ship Software	lines of code	reuse category	Basic COPLIMO		variant lines of code variant cost savings	2.35 MSLOC
ASW Combat System Cross-domain	system component lines of code	reuse category	Basic COPLIMO	Requirements models, OVM	system costs system lines of code	18 system components 2.1 MSLOC
DoN UUV Missions	system requirements system interfaces	reuse category complexity level	COSYSMO 2.0	Requirements models activity models		57 system requirements 14 system interfaces
Mine Counter Measure UUVs	system requirements system interfaces	reuse category complexity level	COSYSMO 2.0	OVM		16 system components

Known cost models were adapted for different system types, processes, and estimation relationships at the systems and software levels. The basic reuse and investment model was supplanted with alternate cost models relevant to the system types under consideration. This was supported by the development of an integrated method for representing architectural variants using orthogonal variability modeling (OVM) to enumerate parametric inputs for COPLIMO.

The rest of this paper will present an overview of the cost modeling followed by a more detailed explanation of the case studies presented in Table 1.

Cost Modeling

The two basic cost models used are the COSYSMO model and the COPLIMO model. The COSYSMO model inputs for system size include requirements and interfaces classified by reuse category and complexity. It uses size weights to account for the relative effort for the reuse categories: New, Designed for Reuse, Modified, Deleted, Adopted, and Managed. The complexity levels also have equivalent size weights for Easy, Nominal, and Difficult ratings.

COPLIMO provides a trade space for determining initial investment and future return on investment (ROI) for product line systems versus non-product line systems. Product line investment models must address two sources of cost investment or savings which were afforded by COSYSMO in this approach. The relative cost of developing product lines is the added effort of developing flexible product line architectures to be most cost-effectively reused across a product line family of applications, relative to the cost of developing a single system. In COSYSMO, this investment cost is captured in the Designed for Reuse category.

The relative cost of reuse is the cost of reusing system architecture in a new product line family application relative to developing new systems. COSYSMO has the categories for Reuse, Modified, Deleted, Adopted, and Managed to quantify the relative costs compared to the New category.

The model size inputs were extracted from the product models for each mission type. Each requirement and interface in the models were further tagged for reuse category and complexity level. The COSYSMO size weights are then applied in the estimation tools.

Model outputs provide decision makers with essential information on product line savings, investment, ROI, cost per mission type, and savings per mission type. It supports the initial investment decision as well as a starting point for planning the individual system developments. The cost and schedule of each system is already estimated and can be planned over time per the mission needs.

Table 18. ROI Analysis for RCDR 1.5 through Six Architecture Alternatives

	Eq. Size - Reuse Model	Non-Reuse Size	PL Effort Savings	Cumulative Savings	ROI	Cumulative ROI
Alternative 1 (Baseline)	23.50	16	-7.50	-7.50	-1.00	-1.00
Alternative 2	3.20	17	13.80	6.30	1.84	0.84
Alternative 3	4.20	18	13.80	20.10	1.84	2.68
Alternative 4	3.80	17	13.20	33.30	1.76	4.44
Alternative 5	4.20	18	13.80	47.10	1.84	6.28
Alternative 6	5.20	19	13.80	60.90	1.84	8.12
PL Reuse Investment	7.5					

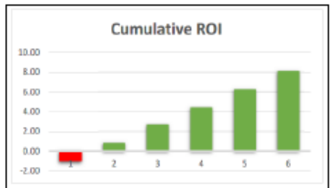


Figure 51. Cumulative ROI for RCDR 1.5

Table 19. ROI Analysis for RCDR 1.6 through Six Architecture Alternatives

	Eq. Size - Reuse Model	Non-Reuse Size	PL Effort Savings	Cumulative Savings	ROI	Cumulative ROI
Alternative 1 (Baseline)	25.00	16	-9.00	-9.00	-1.00	-1.00
Alternative 2	3.20	17	13.80	4.80	1.53	0.53
Alternative 3	4.20	18	13.80	18.60	1.53	2.07
Alternative 4	3.80	17	13.20	31.80	1.47	3.53
Alternative 5	4.20	18	13.80	45.60	1.53	5.07
Alternative 6	5.20	19	13.80	59.40	1.53	6.60
PL Reuse Investment	9.0					

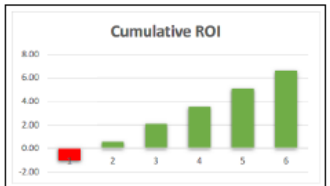


Figure 52. Cumulative ROI for RCDR 1.6

Table 20. ROI Analysis for RCDR 1.8 through Six Architecture Alternatives

	Eq. Size - Reuse Model	Non-Reuse Size	PL Effort Savings	Cumulative Savings	ROI	Cumulative ROI
Alternative 1 (Baseline)	28.00	16	-12.00	-12.00	-1.00	-1.00
Alternative 2	3.20	17	13.80	1.80	1.15	0.15
Alternative 3	4.20	18	13.80	15.60	1.15	1.30
Alternative 4	3.80	17	13.20	28.80	1.10	2.40
Alternative 5	4.20	18	13.80	42.60	1.15	3.55
Alternative 6	5.20	19	13.80	56.40	1.15	4.70
PL Reuse Investment	12.0					

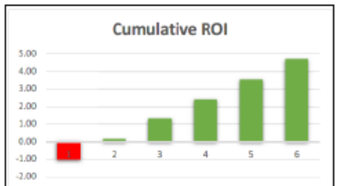


Figure 53. Cumulative ROI for RCDR 1.8

Basic COPLIMO

The basic version of COPLIMO supports software product line cost estimation and ROI analysis within the scope of the product line life cycle. Basic COPLIMO shown in Figure 1 consists of two components:

- Product line development cost model
- Annualized post-development life cycle extension

The model is based on the COCOMO II software cost model and has been statistically calibrated to 161 projects, representing 18 diverse organizations.

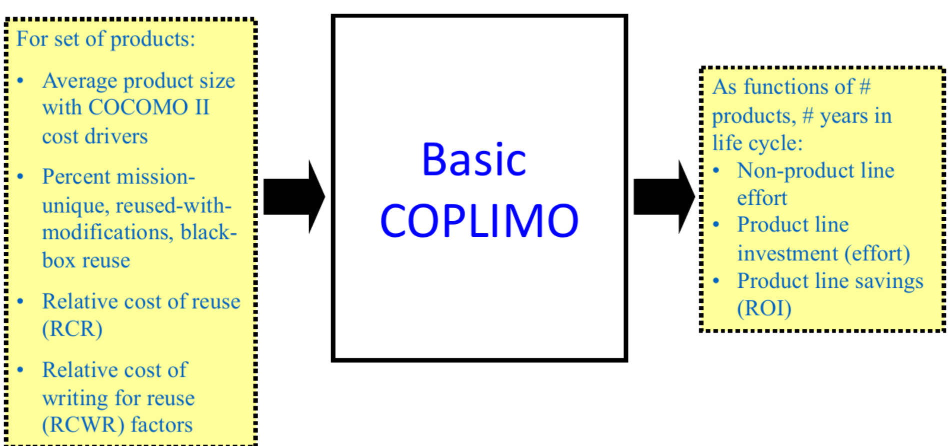


Figure 1. Basic COPLIMO

Table 2 is a list of extensions that have been made to basic COPLIMO.

Table 2. Basic COPLIMO Extensions

- Separate factors for calculating software RCR
  - Design, code, test fractions modified
  - Software understanding, assessment factors
- Separate factors for calculating software RCWR
  - Reusability, reliability, documentation
- Full set of COCOMO II cost drivers
- Maintenance and life cycle cost estimation
- Components with different sizes, RCR, RCWR factors
- Present-value discounting of future savings
- Monte Carlo probability distributions

System Product Line Investment Model

Figure 2 presents the system product line investment model. It differs from the Basic COPLIMO model in that the results are based on the product line total ownership costs and the product line flexibility investment. The model uses generic system components for software and hardware, size-based modeling or direct cost, annual change cost, and full life cycle total ownership cost.

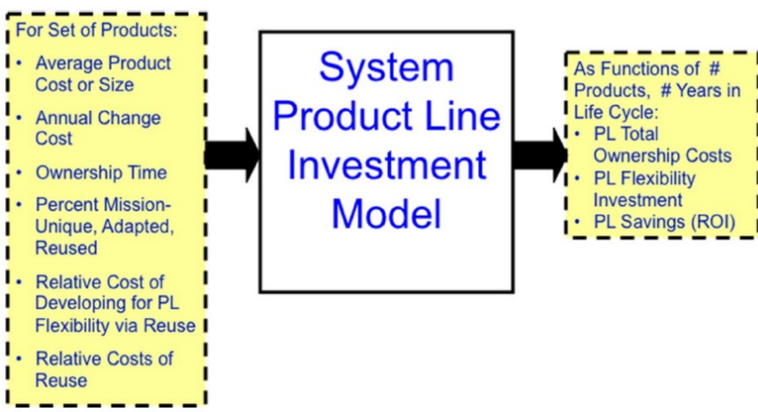


Figure 2. The System Product Line Investment Model

Selected Cost/ROI Modeling Tools

Figure 3 presents selected cost and return on investment modeling tools.

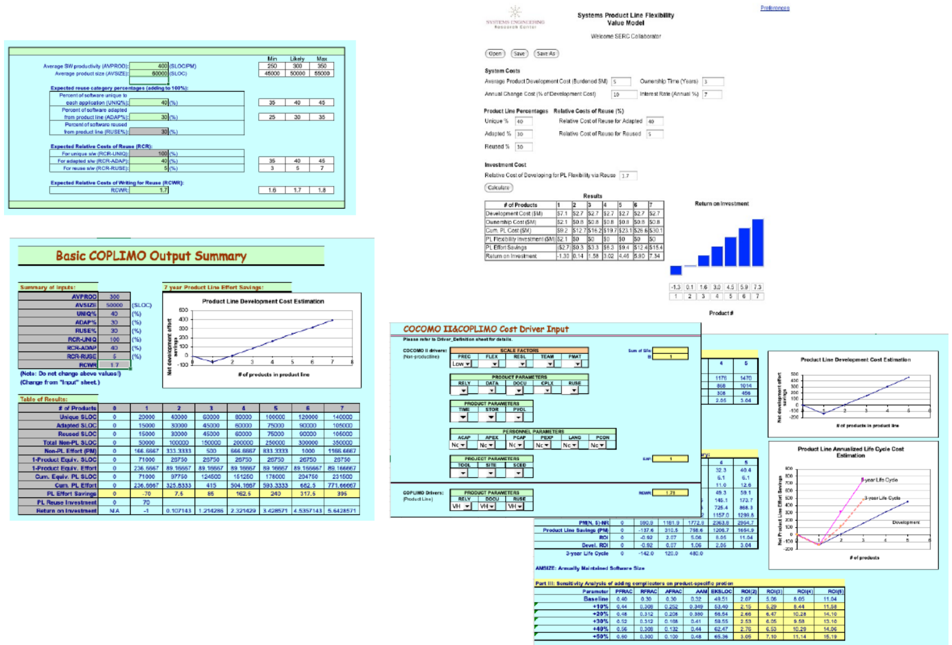


Figure 3. Selected Cost and Return on Investment Modeling Tools

Of interest is the tool in the upper right corner of Figure 3. Known as the Systems Product Line Flexibility Value Model, one can adjust system costs, product line percentages, and the relative cost of reuse to see how they impact ROI. A larger version of Figure 3 is shown in Figure 4.

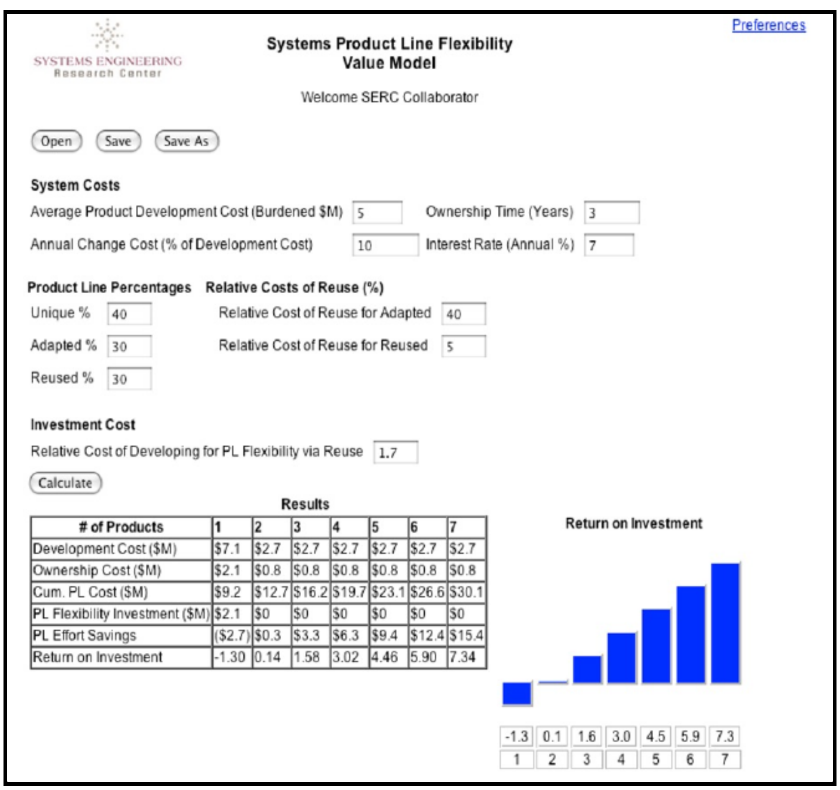


Figure 4. Systems Product Line Flexibility Value Model
(http://coplimo.org/tools/flexibility)

Case Studies

The following section provides an overview of the case studies shown in Table 1. Each of the case studies is readily available through the NPS institutional archive: Calhoun (https://library.nps.edu/nps-archive).

Combat Systems Product Lines

The approach to the initial case study used a domain-specific model-based system engineering (MBSE) framework focused on a reference architecture of a general combat system product line. The MBSE approach was integrated with COPLIMO for size inputs derived from the MBSE models including OVM.

Specifically, the reference architecture was based on an underlying detect-control-engage architecture. This top-level functional architecture was then allocated to mission-specific system components which were assessed for reuse. The OVM model was used to quantify change percentages for new, modified, and deleted components. The method used is described below.

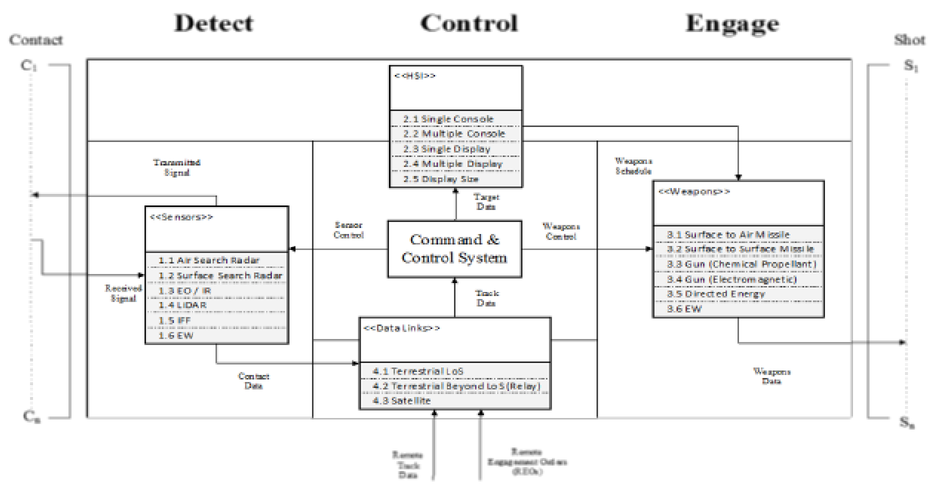


Figure 5. Generic Combat System Reference Architecture

Method Overview

Step 1: Describe a general domain model of the given system with common elements. For a combat system, the architecture includes sensors, weapons, and hardware/software which are formally modeled to identify common functions and variations for different case studies.

Step 2: Develop a reference product architecture with variation points. Variation points are identified for sensors, consoles, weapons, and data links with alternative choices to serve as cost model inputs.

- Map existing systems to the reference architecture
- Collect empirical costs and map them to system elements from above

Empirical cost data from DoD programs is allocated to the system functions in the architecture models to calibrate and populate the cost model for specific system configurations. Collect empirical costs and map them to system elements from above.

Figure 6 is an example OVM from Alves' thesis (Alves, 2022). The actual OVM presents more detail.

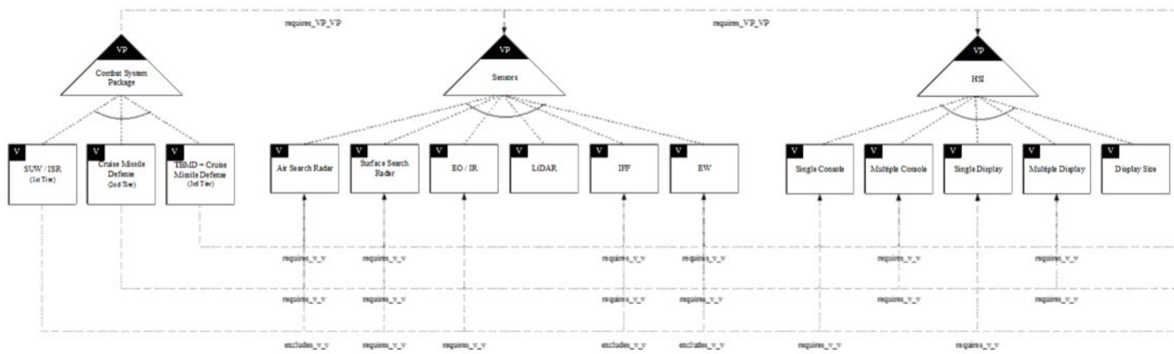


Figure 6. Example Orthogonal Variability Model (Alves, 2022)

Figure 7 is a table of the OVM symbology. Hause describes how the OVM can be used in the block definition diagram (BDD) “to define relationships between and properties of the elements which are represented on those diagrams” (Hause, 2014).

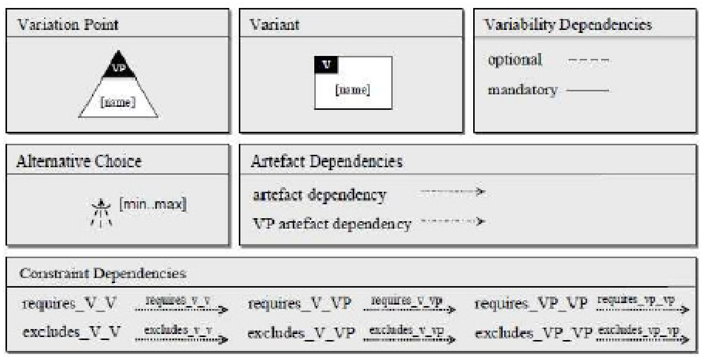


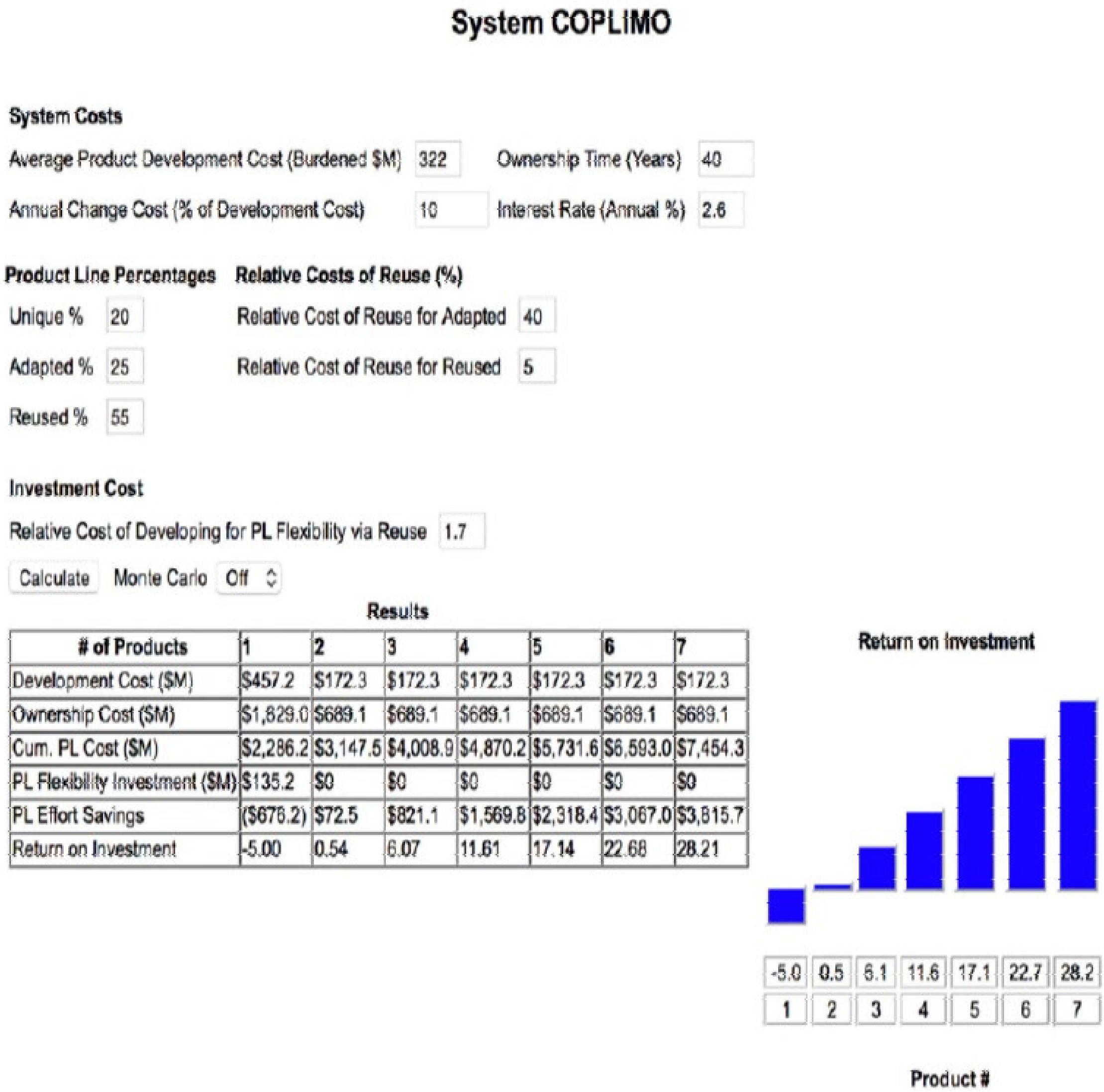
Figure 7. OVM Notation (Pohl et al., 2005)

Table 3 presents example product line components used in the case study.

Table 3. Example Product Line Components

Product Line Classification	Variant	Justification
Variation Point: Sensors		
Adapted	Air Search Radar	Power, beam forming, and search / track functions different for 2nd and 3rd tier packaged variants.
Adapted	EW	Power and physical size requirements may be different for 2nd and 3rd tier packaged variants.
Reused	Surface Search Radar	Physical size and capabilities of sensor can be used for 1st, 2nd, and 3rd tier packaged variants.
Reused	EO / IR Sensor	See Surface Search Radar justification.
Reused	LiDAR	See Surface Search Radar justification.
Reused	IFF	Hardware and interfaces are the same for 2nd and 3rd tier packaged variants.
Variation Point: HSI		
Product Line Classification	**Variant**	**Justification**
Reused	Single Console	Consoles common across 1st, 2nd, and 3rd tier packaged variants.
Reused	Multiple Console	See Single Console justification.
Reused	Single Display	Displays common across 1st, 2nd, and 3rd tier packaged variants.
Reused	Multiple Display	See Single Display justification.
Adapted	Display Size	Displays are common but size can be specified by customer.
Variation Point: Data Links		
Product Line Classification	**Variant**	**Justification**
Reused	Terrestrial LoS	Data links standardized across US and NATO platforms, therefore they will also be common across 1st, 2nd, and 3rd tier packaged variants.
Reused	Terrestrial Beyond LoS	See Terrestrial LoS justification.
Reused	Satellite	See Terrestrial LoS justification.
Variation Point: Weapons		
Product Line Classification	**Variant**	**Justification**
Mission Unique	Surface to Air Missile	Ranges and kill mechanisms are different for 2nd and 3rd tiers.
Mission Unique	Surface to Surface Missile	Ranges and size of missile different for 1st, 2nd and 3rd tiers based on mission and ship size.
Mission Unique	Gun Electro-Magnetic	Power and size constraints dependent on ship size and cost for 2nd and 3rd tiers.

Figure 8 presents example cost and ROI results for the cruise missile product line included in the case study.



# of Products	1	2	3	4	5	6	7
Development Cost ($M)	$457.2	$172.3	$172.3	$172.3	$172.3	$172.3	$172.3
Ownership Cost ($M)	$1,829.0	$689.1	$689.1	$689.1	$689.1	$689.1	$689.1
Cum. PL Cost ($M)	$2,286.2	$3,147.5	$4,008.9	$4,870.2	$5,731.6	$6,593.0	$7,454.3
PL Flexibility Investment ($M)	$135.2	$0	$0	$0	$0	$0	$0
PL Effort Savings	($676.2)	$72.5	$821.1	$1,569.8	$2,318.4	$3,067.0	$3,815.7
Return on Investment	-5.00	0.54	6.07	11.61	17.14	22.68	28.21

Figure 8. Results for Cruise Missile Product Line (Chance, 2019)

Table 4 presents the detailed COPLIMO model for the Aegis combat system as analyzed. Actual values were used in the analysis. However, the values in Table 4 are representative sizes per agreement with Lockheed Martin.

Table 4. AEGIS Combat System Detailed COPLIMO*

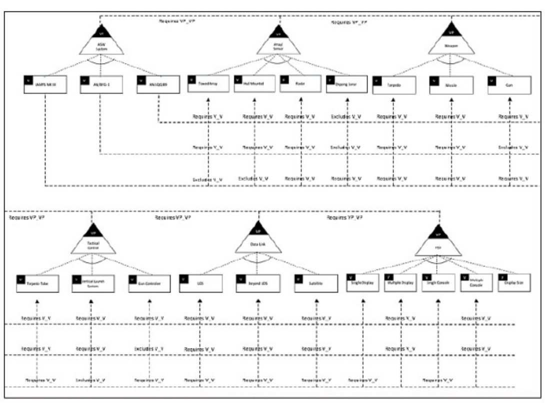
# of Products	0	1	2	3	4	5	6
Aegis Baseline		A	B	C	D	E	F*
Unique SLOC	0	100000	200000	400000	200000	400000	200000
Adapted SLOC	0	540000	570000	630000	750000	810000	930000
Reused SLOC	0	1260000	1330000	1470000	1750000	1890000	2170000
Total Non-PL SLOC	0	1900000	4000000	6500000	9200000	12300000	15600000
Total Non-PL Effort (PM)	0	11656	24540	39877	56442	75460	95706
1-Product ESLOC	0	3304000	379000	379000	379000	379000	379000
1-Product Equiv. Effort (PM)	0	20270	2325	2325	2325	2325	2325
Cum. ESLOC	0	3304000	3683000	4062000	4441000	4820000	5199000
Cum. PL Effort (PM)	0	20270	22595	24920	27245	29571	31896
PL Effort Savings (PM)	0	-8613	1945	14957	29196	45890	63810
PL Reuse Investment	0	8613	0	0	0	0	0
Per Product Non PL Effort (PM)	0	11656	12883	15337	16564	19018	20245
Per Product PL Effort (PM)	0	2325	3034	4451	3604	5021	4175
Per Product Cost Savings (PM)	0	-8613	9850	10887	12960	13997	16071
Per Product Cost Avoidance	0	-173.89%	23.55%	29.02%	21.76%	26.40%	20.62%
Cum. ROI	0	-2.00	-0.86	0.41	1.91	3.54	5.40

*Actuals were used but these are representative sizes

Cross-Domain Antisubmarine Warfare Combat Systems

This case study investigated the application of a product line model to both surface ship and submarine combat systems. Most of the software performs the same function regardless of whether the antisubmarine warfare (ASW) combat system is aboard a surface ship or a submarine. The variability is in the sensors and weapons. Current acquisition practice is to procure the ASW combat system separately from different sources thus there is little reuse, if any.

- Product line orthogonal variability model



- Variant Classification

Variation Point: Array/Sensor		
Product Line Classification	Variant	Supporting Rationale
Adapted	Towed Array	Array only used for surface and subsurface platforms
Adapted	Hull Mounted Array	Array only used for surface and subsurface platforms
Reused	Radar	Standard across all systems
Unique	Dipping SONAR	SONAR only used for LAMPS MK III
Variation Point: Weapon		
Product Line Classification	Variant	Supporting Rationale
Unique	Torpedo	Weapon dependent on ship size and mission
Unique	Missile	Weapon dependent on ship size and mission
Adapted	Gun	Size of the gun varies between air and surface ship ASW systems
Variation Point: Tactical Control		
Product Line Classification	Variant	Supporting Rationale
Adapted	Torpedo Tube	Torpedo tube varies between systems
Adapted	Vertical Launch System	Vertical launch system varies between systems
Adapted	Gun Controller	Guns vary between systems
Variation Point: Data Link		
Product Line Classification	Variant	Supporting Rationale
Reuse	LOS	Standard across all systems
Reuse	Beyond LOS	Standard across all systems
Reuse	Satellite	Standard across all systems
Variation Point: HSI		
Product Line Classification	Variant	Supporting Rationale
Reuse	Single Display	Displays common across systems
Reuse	Multiple Display	Displays common across systems
Reuse	Single Controller	Controllers common across systems
Reuse	Multiple Controller	Controllers common across systems
Adapted	Display Size	Size specified by restrictions of the system

Figure 9. ASW Product Line Orthogonal Variability Model (Fraine et al., 2019)

Table 5 presents the results of a most likely scenario where the ASW combat system was built as a product line. The net development effort savings follows the typical path for product line development.

Table 5. Cross-Domain ASW Combat System Product Line Most Likely Scenario (Fraine et al., 2019)

Summary of Inputs				
	Baseline Product	Product 1 (LAMPS MK III)	Product 2 (AN/BYG-1)	Product 3 (SQQ-89)
Product Cost ($M)	$221.30	$221.30	$221.30	$221.30
Ownership Time (years)	40	25	40	40
Annual Change Cost (%)	10	10	10	10
Interest Rate (%)	3.63	3.63	3.63	3.36
Unique (%)	11.1	13.3	13.3	12
Developed for Product Line Reuse (%)	88.8	0	0	0
Adapted (%)	0	33.3	33.3	41
Reused (%)	0	53.3	53.3	47
RCR-Unique (%)	100	100	100	100
RCR-Adapted (%)	40	40	40	40
RCR-Reused (%)	5	5	5	5
RCWR	1.7	1.7	1.7	1.7
Non-Product Line Cost ($M)		189	212.5	262.4

Summary of results						
Product #	Development Cost ($M)	Ownership Cost ($M)	Investment	Savings ($M)	Cost Avoidance	Cum. ROI
0	$358.64	1434.56	$137.56	($687.80)		-6.00
1	$64.81	162.02	0	$546.95	85%	-2.02
2	$64.81	259.23	0	$781.35	83%	3.66
3	$97.74	390.98	0	$766.25	68%	9.23

Unmanned Underwater Vehicles Product Lines

The U.S. Navy has nine primary missions:

1. Intelligence, surveillance, and reconnaissance (ISR),
2. Mine countermeasures (MCM),
3. Antisubmarine warfare (ASW),
4. Inspection and identification (INID),
5. Oceanography (OO),
6. Communication or navigation network node (CN3),
7. Payload delivery (PD),
8. Information operations (IO), and
9. Time critical strike (TCS).

Detailed analyses for the UUV mission types were used to develop the SysML models that encapsulated system size and complexity measures. Analysis and comparison of the defined UUV missions identified ISR as having the most commonality across the set and was chosen as the reference architecture. Development of the ISR UUV constituted the investment costs.

Requirements models were generated and provided enumeration of system requirements by reuse type and complexity. Detailed executable activity models of mission operations were used to quantify interfaces with their complexities for inputs to the cost models.

Figure 10 illustrates the relationship between the mission sets and the COPLIMO model. The model is extended further by the use of the COSYMO 2.0 model

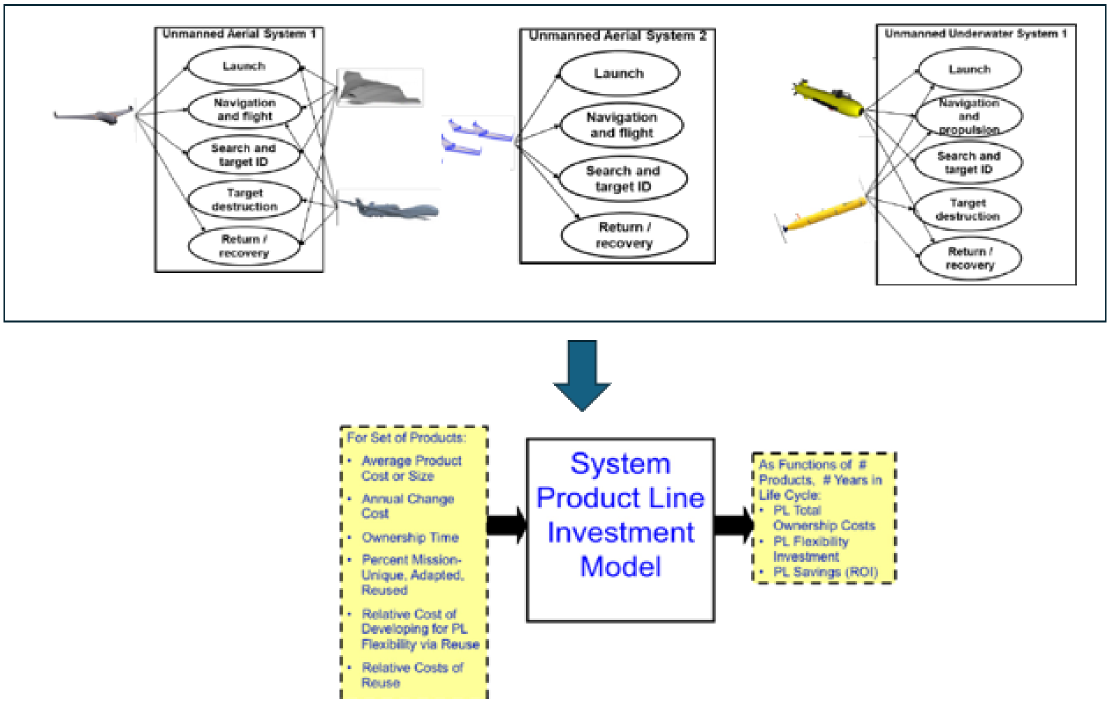
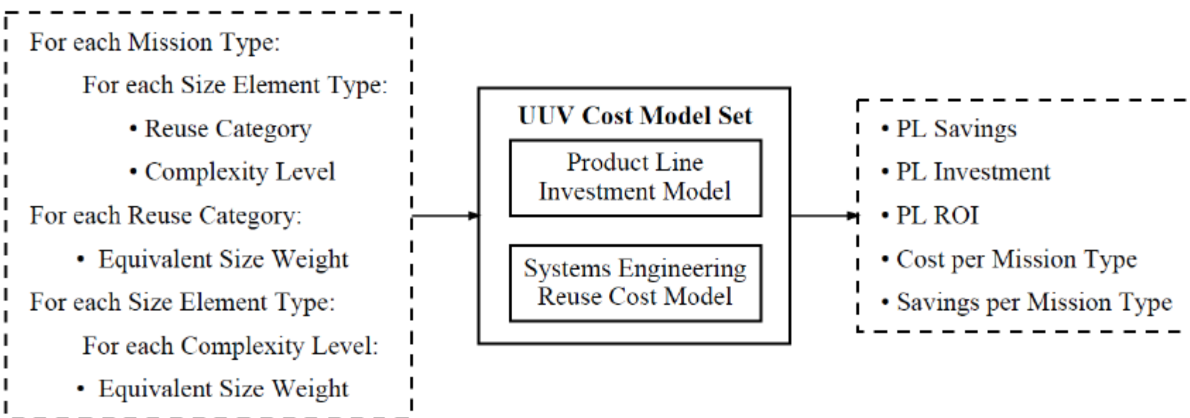


Figure 10. Example Unmanned Systems Product Line Commonality

Figure 11 is Figure 10 extended for the UUV mission set. Table 6 lists the reuse categories that satisfy the UUV requirements.



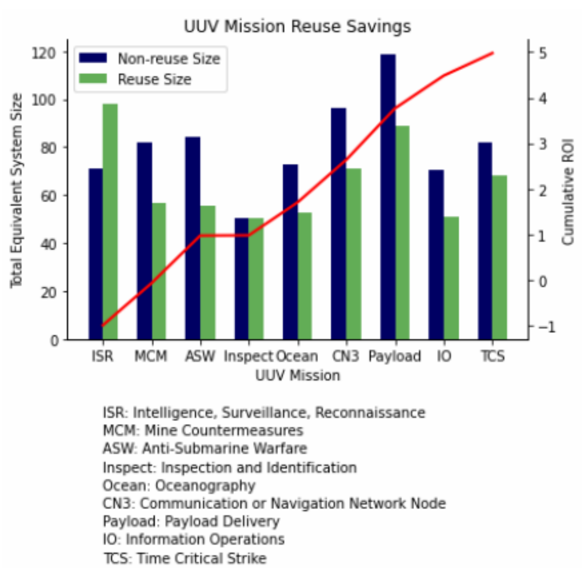
Where:
Size Element Types = (Requirements, Interface, Algorithms, Scenarios)
Reuse Categories = (New, Designed for Reuse, Modified, Deleted, Adopted, Managed)
Complexity Levels = (Easy, Nominal, Difficult)

Figure 11. COPLIMO Extended with COSYSMO 2.0 for UUV Missions

Table 6. COSYSMO 2.0 Reuse Categories Interpreted for UUV Requirements

Category	Definition for Requirements	Definition for Interfaces	Weight
New	Similar requirement does not exist in the baseline (completely new)	Similar interface does not exist in the baseline (completely new)	1.00
Designed for Reuse	New requirement and includes extra investment to enable potential reusability	New interface and includes extra investment to enable potential reusability	1.38
Modified	Change to the requirement's Measures of Effectiveness (MOEs)	Interface is tailored to the mission	0.65
Deleted	Similar requirement does not exist in new system	Similar interface does not exist in new system	0.51
Adopted	Change to the requirement's Measures of Performance (MOPs)	Interface is incorporated unmodified with testing	0.43
Managed	Requirement does not change from the baseline	Interface is incorporated unmodified with minimal testing	0.15

Figure 12 presents the UV mission reuse savings and ROI. The figure also list some of the important considerations involved in building the model.



- Requirements and interfaces from UUV MBSE models were enumerated and input into the COSYSMO cost model.
- This indicator displays the total equivalent system sizes and resultant ROI of a product line approach for UUV systems with overlapping mission capabilities.
- The savings for subsequent missions are the differences between a traditional non-reuse approach and the product line reuse approach.
- The cumulative ROI is the net savings over
- time divided by the investment cost based on the relative sizes.
- The size is used as input to systems engineering cost models to quantify estimated costs.
- The equivalent size difference represents a work savings, and added equivalent size represents the additional work investment to make the UUV baseline reusable.

Figure 12. UUV Mission Reuse Savings and ROI (Haller et al., 2022)

Mine Counter Measure UUV Product Line Modeling

The most mature case study to date is the Mine Counter Measure (MCM) UUV study completed by Alves (Alves, 2022). This study will provide a foundation for work going forward.

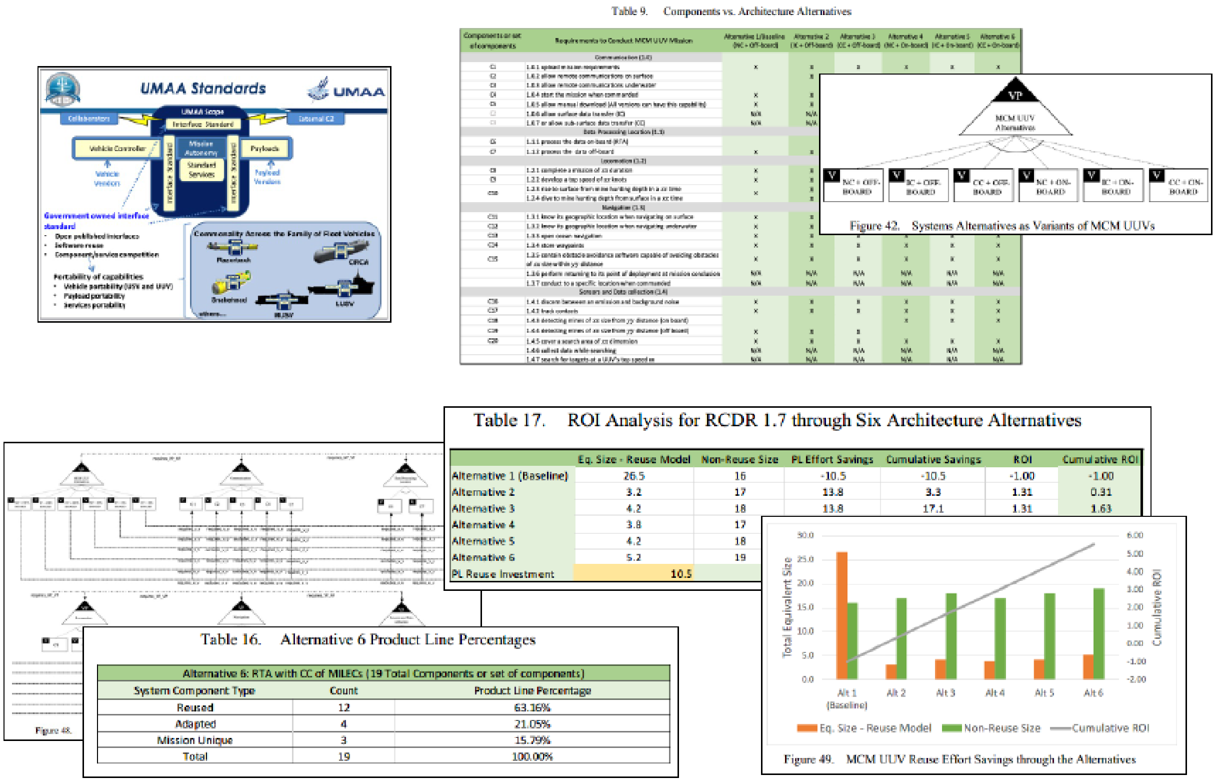


Figure 13. A Montage of the Analysis Steps Involved in the MCM UUV Product Line Modeling (Alves, 2022)

MCM UUV Economic Sensitivity Analysis

Table 18. ROI Analysis for RCDR 1.5 through Six Architecture Alternatives

	Eq. Size - Reuse Model	Non-Reuse Size	PL Effort Savings	Cumulative Savings	ROI	Cumulative ROI
Alternative 1 (Baseline)	23.50	16	-7.50	-7.50	-1.00	-1.00
Alternative 2	3.20	17	13.80	6.30	1.84	0.84
Alternative 3	4.20	18	13.80	20.10	1.84	2.68
Alternative 4	3.80	17	13.20	33.30	1.76	4.44
Alternative 5	4.20	18	13.80	47.10	1.84	6.28
Alternative 6	5.20	19	13.80	60.90	1.84	8.12
PL Reuse Investment	7.5					

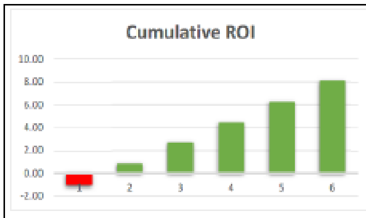


Figure 51. Cumulative ROI for RCDR 1.5

Table 19. ROI Analysis for RCDR 1.6 through Six Architecture Alternatives

	Eq. Size - Reuse Model	Non-Reuse Size	PL Effort Savings	Cumulative Savings	ROI	Cumulative ROI
Alternative 1 (Baseline)	25.00	16	-9.00	-9.00	-1.00	-1.00
Alternative 2	3.20	17	13.80	4.80	1.53	0.53
Alternative 3	4.20	18	13.80	18.60	1.53	2.07
Alternative 4	3.80	17	13.20	31.80	1.47	3.53
Alternative 5	4.20	18	13.80	45.60	1.53	5.07
Alternative 6	5.20	19	13.80	59.40	1.53	6.60
PL Reuse Investment	9.0					

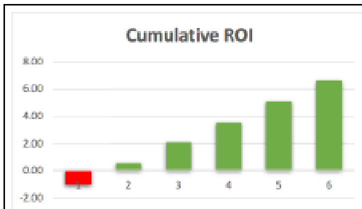


Figure 52. Cumulative ROI for RCDR 1.6

Table 20. ROI Analysis for RCDR 1.8 through Six Architecture Alternatives

	Eq. Size - Reuse Model	Non-Reuse Size	PL Effort Savings	Cumulative Savings	ROI	Cumulative ROI
Alternative 1 (Baseline)	28.00	16	-12.00	-12.00	-1.00	-1.00
Alternative 2	3.20	17	13.80	1.80	1.15	0.15
Alternative 3	4.20	18	13.80	15.60	1.15	1.30
Alternative 4	3.80	17	13.20	28.80	1.10	2.40
Alternative 5	4.20	18	13.80	42.60	1.15	3.55
Alternative 6	5.20	19	13.80	56.40	1.15	4.70
PL Reuse Investment	12.0					

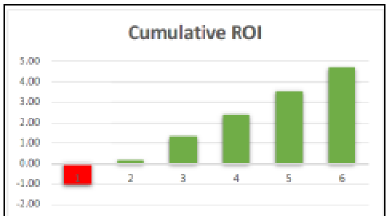


Figure 53. Cumulative ROI for RCDR 1.8

Figure 14. A Sensitivity Analysis of ROI for Architecture Alternatives (Alves, 2022)

The case study outcome was a substantial ROI of five for the product line approach over the single system approach for the nine UUV systems. This result corroborates previous product line economic analyses, demonstrating that many DoD systems and other types of system families would benefit from a product line strategy.

Conclusions and Future Work

COPLIMO provides a useful trade space for determining initial investment and future ROI with respect to product line systems versus non-product line systems.

- Virtually all case studies have demonstrated high ROI of product line practices on defined DoD missions.
- System architectures for chosen domains should focus on the product line, instead of mission specific systems. Plan for the reuse of system components over time.
- Applying the engineering product line methodology to system architecture design and development needs to happen at the earliest stage of design.
- System architectures for unmanned systems should focus on the product line, instead of mission specific systems. The product line modeling approach has a broader application for acquiring systems that are based on similar functions and will be applied to future case studies.

Future work includes additional case studies and combined modeling of systems effectiveness with the economics of product lines. The model integration is being further streamlined. We are developing improved tools for SysML 2 to automate the product and cost model integration. With this, we can also include a broader range of system size information from activity models, use case models and sequence models.

References

Alves, F. (2022, September). *Economic tradeoff analysis of a product line architecture approach through model-based systems engineering: A case study of future mine countermeasures unmanned underwater vehicles* [Master's thesis, Naval Postgraduate School].

Boehm, B., Brown, A. W., Madachy, R., & Yang, Y. (2004). A software product line life cycle cost estimation model. *Proceedings of the 2004 International Symposium on Empirical Software Engineering*, ISESE'04, 156–164.

Chance, K. (2019, June). *Naval combat systems product line economics: Extending the constructive product line investment model for the Aegis combat system* [Master's thesis, Naval Postgraduate School].

Fraine, N. D., Jackson-Henderson, T., & Manfredo, V. (2019, June). *Naval ASW combat system product line architecture economics* [Capstone project, Naval Postgraduate School].

Hall, R. (2018, June). *Utilizing a model-based systems engineering approach to develop a combat system product line* [Master's thesis, Naval Postgraduate School].

Haller, K., Kolber, D., Storms, T., Weeks, J., & Weers, W. (2022, March). *Unmanned underwater vehicle mission systems engineering product reuse return on investment* [Capstone project, Naval Postgraduate School].

Hause, M. (2014, August). Model based product line engineering enabling product families with variants. *Proceedings of the 2014 NDIA Ground Vehicle Systems Engineering and Technology Symposium*.

[4] Madachy, R., & Green, J. (2019, May). Naval combat system product line architecture economics. *Proceedings of the 16th Annual Acquisition Research Symposium*.

Madachy, R., & Green, J. (2021, February). *Naval combat system product line architecture economics*. NPS Acquisition Research Program.

Pohl, K., Bockle, G., & Lindon, F. (2005). *Software product line engineering: Foundations, principles, and techniques.* Springer.

Systems Engineering Research Center. (2010). *Valuing flexibility for complex engineering systems RT-18 final report* (Report No. SERC-2010-TR-010-1).

Next Generation Logistics Ship Automation and Uncrewed Underway Replenishment

Dr. Thomas Housel—Professor of Information Science, Naval Postgraduate School.

Dr. Johnathan Mun—Professor of Research, Naval Postgraduate School

Dr. Shelley Gallup—Associate Professor of Research, Naval Postgraduate School

LCDR Nabil H. Tahan (MSC, USN)—Doctoral Candidate

Abstract

The objective of this research is to analyze and propose comprehensive, adaptive, reusable, and innovative modeling processes to assist the U.S. Navy (USN) and Department of Defense (DoD) with Next Generation Logistics Ship (NGLS) automation for estimating and modeling underway replenishment (UNREP). The processes aim to assist in calculating, modeling, valuing, and optimizing a framework for estimating and modeling the UNREP equipment and emergency breakaway investigations. With a heightened strategic focus on distributed maritime and expeditionary advanced operations, this research suggests advanced technological research and development of unmanned UNREP equipment and support systems focused on uncrewed or minimally crewed combat logistics support operations.

The considerable rise in unmanned surface vessels, as well as automated and uncrewed UNREP research, demands the exploration of fueling and storage replenishment at sea, cargo stowage, and handling for unmanned or minimally staffed UNREP equipment operation. Technology development durations, investment costs, and benefits are calculated via this research, and ways to identify and execute emergency breakaway uncrewed or minimally crewed approaches are proposed. This research uses advanced analytical integrated risk management methodologies such as Monte Carlo simulation, stochastic forecasting for uncertainty modeling, knowledge value add analysis, portfolio optimization, prototype strategic flexibility alternatives, technology development durations, and investment options. Based on the results, the research anticipates substantial performance improvements alongside the UNREP processes category, aided by potential new automation.

This research proposes possible paths for advanced technology research and development of unmanned UNREP equipment and support systems, using advanced analytical techniques for technology development timelines estimations, investment analysis, and portfolio optimization. By outlining various strategies and options for identifying situations that warrant standard or emergency procedures, this research informs decision-makers on optimal conditions for UNREP using appropriate technology options that would minimize cost and maximize the efficiency and effectiveness of uncrewed or minimally crewed operations.

Introduction

Research by Koteskey (2020) highlighted the significant portion of refueling time that in-port refueling takes up for U.S. naval surface combatants, despite the Navy consistently deploying around 100 ships post–Cold War. In a major conflict, the inability to access conventional supply ports may make underway replenishment (UNREP) crucial for delivering necessary equipment, supplies, and forces during initial operational phases. To address this need, the research proposes exploring automated and uncrewed UNREP options, particularly with the emergence of unmanned surface vessels (Military Analysis Network, 1999). The study aims to identify potential approaches for unmanned or minimally staffed operation of UNREP equipment, such as fueling and storing replenishment at sea, cargo stowage, and handling. It also outlines paths for advanced technology research and development of unmanned UNREP equipment and support systems, estimating technology development timelines, required investments, and potential benefits.

Additionally, the investigation focuses on emergency procedures for UNREP, including emergency breakaways, and identify approaches for identifying situations warranting emergency breakaways and executing them using uncrewed or minimally crewed methods. The research estimates technology development timelines and required investments for these emergency protocols. Furthermore, the project aims to enhance decision-making processes by introducing advanced analytically based methods of integrated risk management, such as Monte Carlo simulation, stochastic forecasting, portfolio optimization, and strategic flexibility options for prototypes. These methodologies are crucial for avoiding costly disasters in system development and will be instrumental in assessing risks and uncertainties associated with UNREP operations. Moreover, the research project analyzes the application of Integrated Risk Management methodology for decision support, incorporating real options valuation methodology. This approach, successfully used in commercial industries and Navy shipbuilding, assesses the total future value of decisions made under high uncertainty. It includes options like waiting, switching, expanding, sequential development, and prototyping to make informed decisions in uncertain environments.

Research Objective

The Next Generation Logistics Ship (NGLS) analysis of alternatives (AOA) guidance directs the Navy to explore a purpose-built uncrewed or optionally crewed solution for NGLS. Across the Department of Defense (DoD), interest remains strong for uncrewed capabilities, with a heightened focus on uncrewed or minimally crewed solutions for the Combat Logistics Force based on the support of Distributed Maritime Operations (DMO) and Expeditionary Advance Base Operations (EABO) requirements. The Navy's study team is developing concepts to support this AOA alternative (DoN, 2001). This task should explore the systems and technologies required to significantly reduce, if not eliminate, the personnel required to conduct UNREP of U.S. Navy ships using U.S. Naval Auxiliary CLF ships.

Literature Survey

Miller et al. (1987) explained UNREP, which is the safe delivery of maximum cargo in the shortest time. The U.S. Navy operates an unusual cargo ship fleet that must unload goods in motion and transfer them to combatant ships only 150 feet apart. The current system is a combination of seamanship and engineering, with the modern UNREP fleet spanning from small breakbulk cargo ships to big, fast, technologically advanced vessels that can concurrently transfer fuel, ammunition, and supplies day or night, in good or bad weather.

The contents of Naval Warfare Publication 4-01.4, *Underway Replenishment*, are intended to prepare both the replenishment ship and the clientship for a replenishment evolution. Specific criteria are created for all replenishment elements at sea, including rig composition, required equipment, and standard operating procedures. UNREP transfers fuel, ammunition, supplies, and troops from one ship to another while the ships are in motion. There are two primary types of UNREP: linked (CONREP) and vertical (VERTREP), and they may be used separately or simultaneously. In connected replenishment, two or more ships steam alongside one another as hoses and lines transfer fuel, ammunition, supplies, and troops (Chen & Fang, 2001).

Several variables favor ship-to-ship refueling over replenishing at a distance. Refueling alongside allows the oiler or other auxiliary ship to simultaneously service two ships, each with several replenishment stations. Additionally, replenishing alongside allows the entire formation of ships to sustain a higher speed (up to 16 knots instead of the 7 to 8 knot maximum for astern refueling). With new innovative technologies, replenishment processes at sea continue to develop. The Standard Tensioned Replenishment Alongside Method (STREAM) is used in both replenishment at sea (RAS) and fueling at sea (FAS) evolutions and is favored over other connected techniques because it allows for greater ship separation. Hewgley and Yakimenko (2009) examined more unorthodox UNREP techniques, such as precision-guided airdrops, which deliver autonomous cargo packages that guide themselves to a precise landing spot by controlling an attached aerodynamic decelerator (parafoil or parachute). The study highlighted the difficulty of resupplying navy warships at sea and the potential benefits of precision airdrop application.

Analyzing UNREP Processes

Curtin's 2001 study developed a conceptual model that combines current UNREP processes with operational scenarios, including Operational Maneuver From the Sea (OMFTS) and Sea Based Logistics (SBL). The model was modified to estimate UNREP cycle times under different conditions. Experiments showed that increasing helicopter lift capacity significantly reduced cycle time, particularly UNREP cycle time. The simulation model also detected constraining resources on crucial operations paths. The findings contribute to the future configuration of amphibious ships for SBL. The study concluded with the following recommendations:

(i) Increasing the lift capacity of the helicopter conducting the UNREP in our model significantly reduced the total cycle time and UNREP cycle time of the inter/intra ship materiel movement process.

(ii) Increasing the number of helicopters used to conduct UNREP operations in our model only marginally decreased the total cycle time. This suggested that the cargo elevators were a bottleneck and could not move the pallets to their final destination at the same rate they received on the LHD.

(iii) Simulation models are used to understand how systems behave. This is especially useful because any "what-if" type experiments can be performed to gain insight into the future system's performance (Kyrkjebø et al., 2004).

UNREP Process

Rowling et al. (2019) developed a scheduling model for indefinitely replenishing a naval task group by UNREP vessels, which had to travel back and forth to a distant port. Constrained optimization can be applied to optimize various measures of effectiveness, such as the amount of slack time a replenishment ship is not needed and the maximum distance offshore a combination of replenishment ships can support naval and joint operations. Dun (1992) examined situations where operational requirements limit the time spent conducting replenishment, focusing on the time needed to transit between ships and the combat value added to the battle group by replenishment. However, the UNREP process has significant risks, including hydrodynamic problems between two moving ships in waves and interaction disturbances in surface vessel maneuvering and position tracking that often require high-precision controllers (Fu & Haddad, 2003). The baseline procedure for the overall UNREP process is enumerated in more detail, providing context for the more general subprocesses described in the Knowledge Value Added (KVA) analysis.

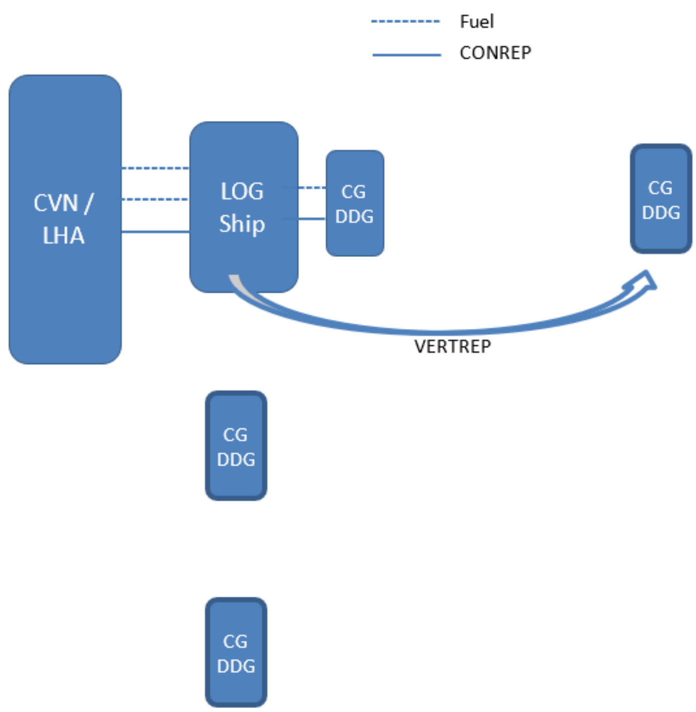


Figure 1. Underway Replenishment

Baseline Procedures

UNREP is broken down into the following modes:

- CONREP
    - Palletized cargo transfer via span wire
    - Personnel transfer via span wire
- FAS
    - Liquid fuel transfer via span wire
- VERTREP
    - Cargo and personnel transfer via helicopter

CONREP and FAS are similar in processes. VERTREP has its own process.

- When we UNREP a carrier strike group (CSG) or amphibious ready group (ARG), we generally work in the formation depicted in Figure 1. A large deck comes alongside the port first, and then a CG/DDG comes alongside the starboard of the LOG. We can do 4 to 5 CG/DDG simultaneously with the large deck.
- When a CG/DDG is complete, they push out to starboard, and we continue to send cargo via helo until complete. The next CG/DDG drifts out to line up for the approach to starboard and then runs the approach. Once the CG/DDG is complete with UNREP, the tactical command (OTC) officer maneuvers them to a screen position.

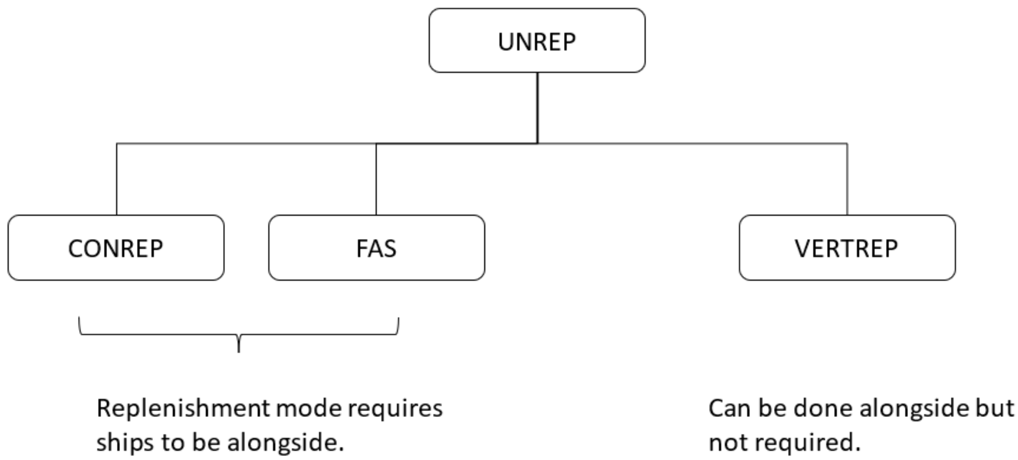


Figure 2. Subprocesses in UNREP

Ships logistic demand signal is sent via the following:

1. **OPSTAT RASREQ:** signals RAS requirements to the OTC and/or the CLF ship
2. **OPTASK RAS:** the OTC/URG commander's or replenishment ship's confirmation message containing details of the UNREP event
3. **OPSTAT RAS:** UNREP ship provides details of rigs and stations
4. **OPSTAT CARGO:** CLF ship report of major cargo remaining after UNREP

Provisions (all but Class III and V supplies, in slide notes [United States Department of the Army, 2008], see appendix) are procured, packed, and palletized at logistics centers via the Naval Sustainment System-Supply (NSS-S). Pallets are sent to sea via single-product ships (oilers [T-AO], dry cargo, and ammo transports [T-AKE]). Cargo is sometimes consolidated into a multiproduct ship (oiler/ammo/dry cargo), which are fast combat logistics ships (T-AOE) that can keep up with the speed of the CSG.

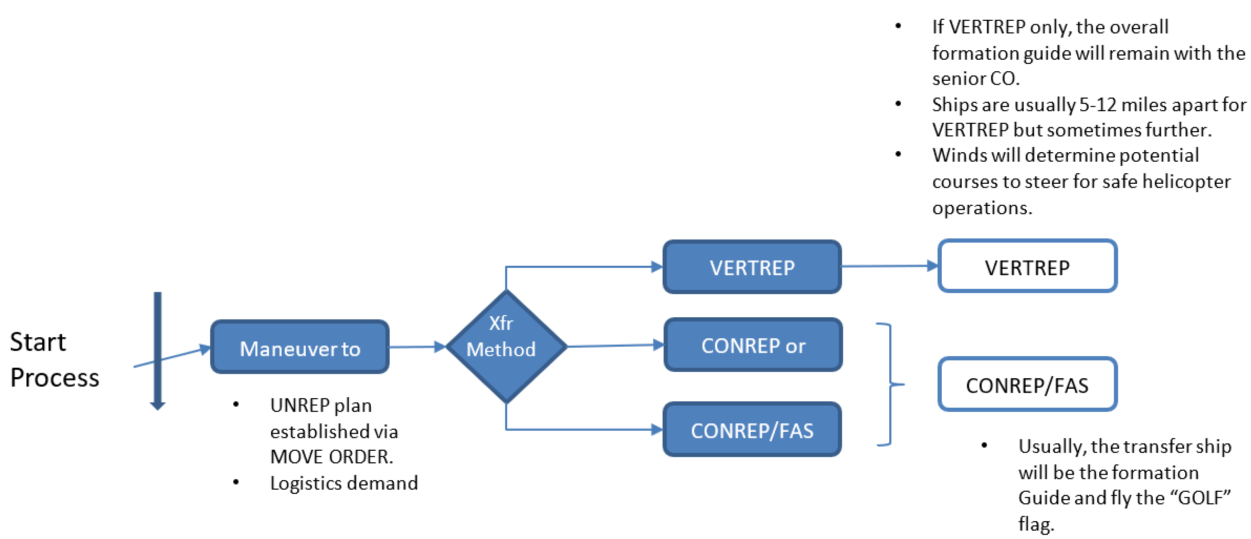


Figure 3. Overall UNREP Process

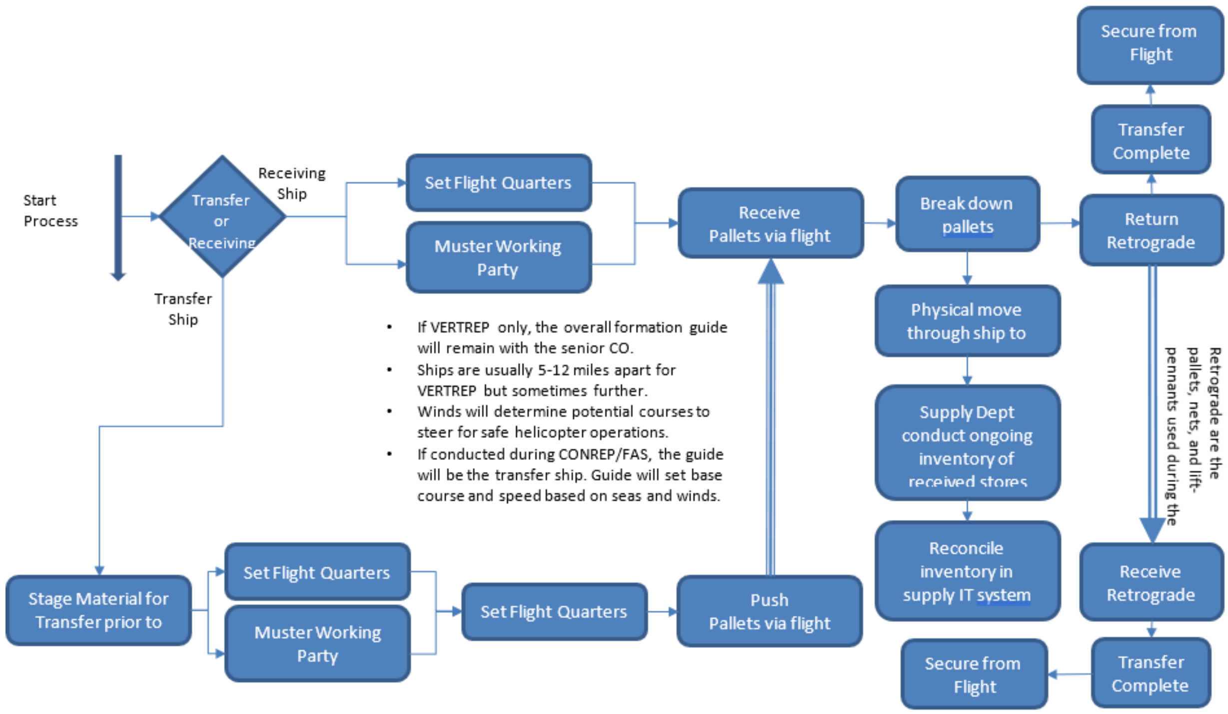


Figure 4. VERTREP Subprocess

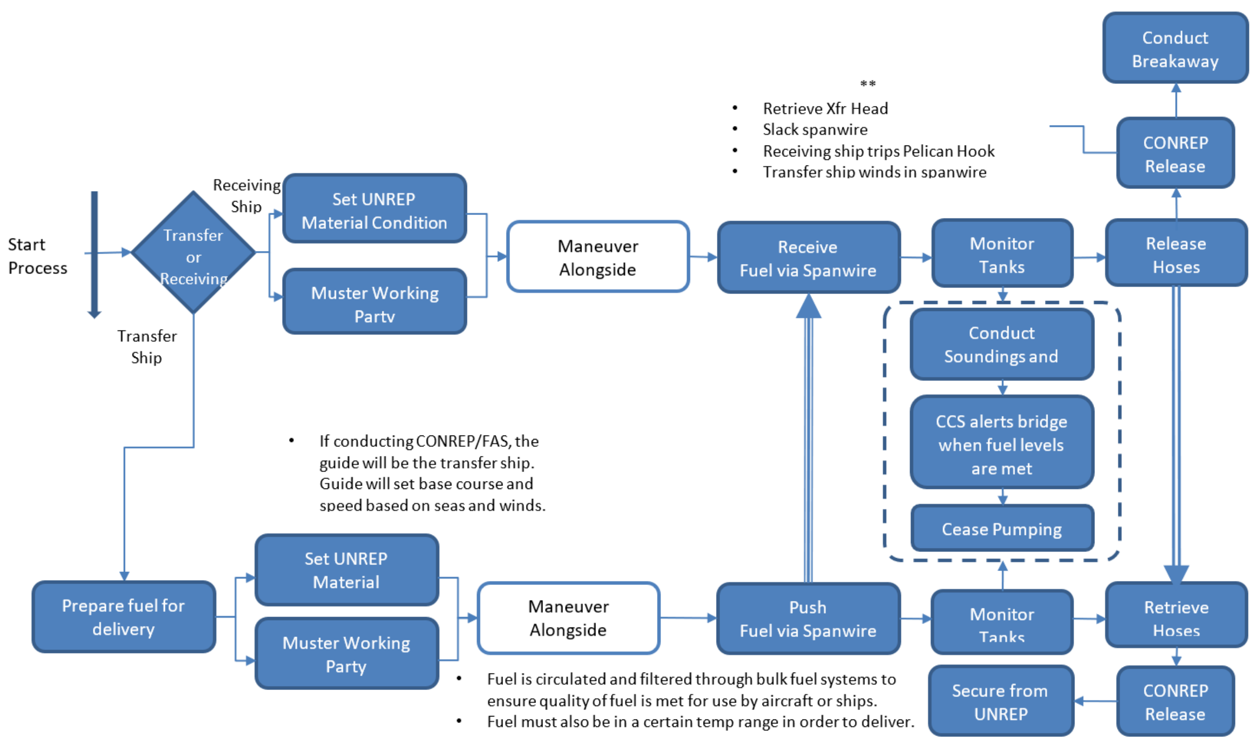


Figure 5. FAS Subprocess

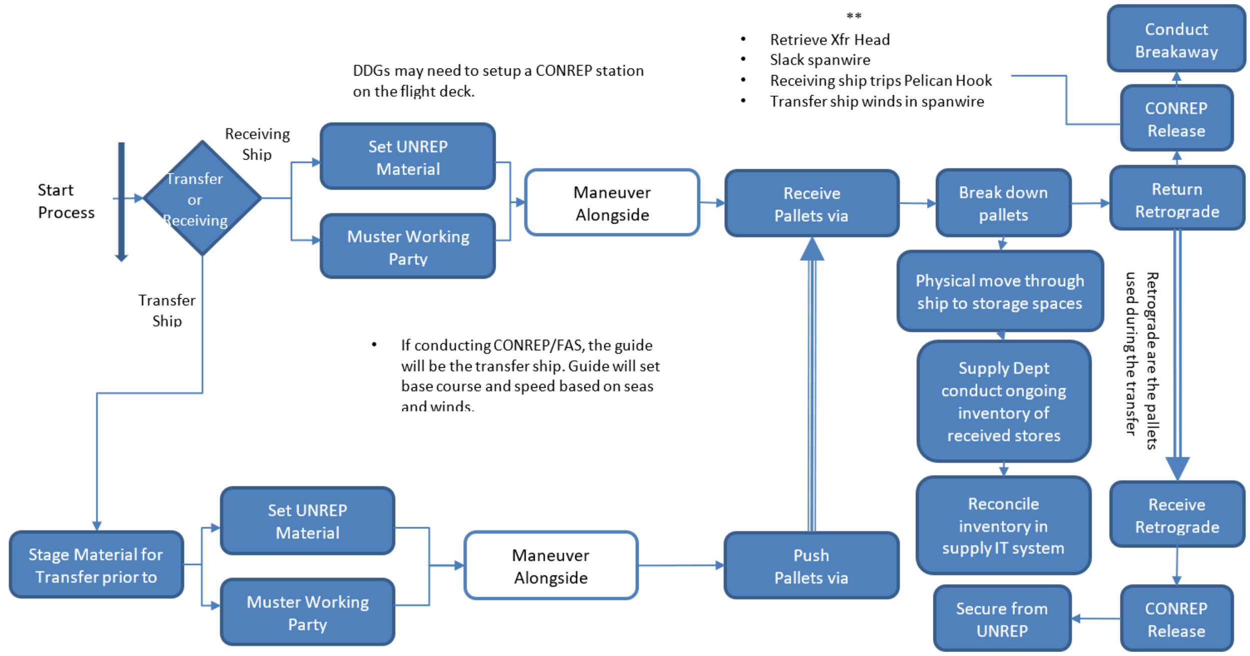


Figure 6. CONREP Subprocess

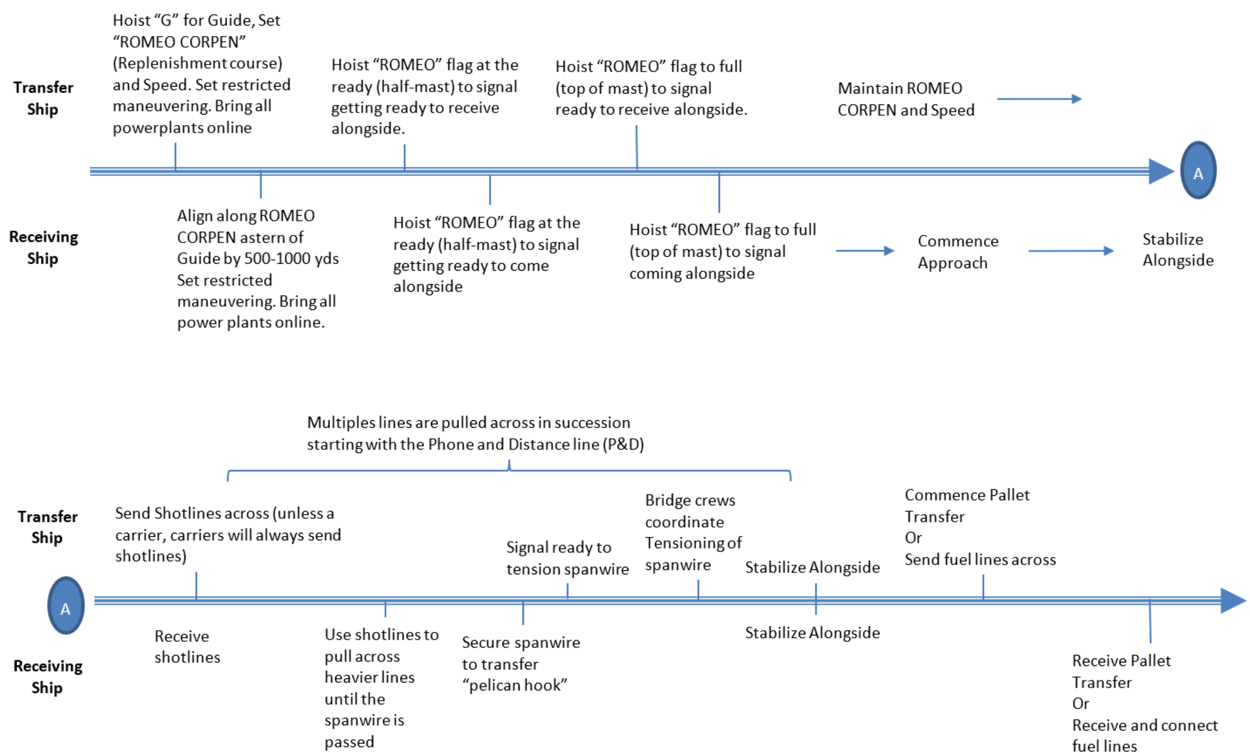


Figure 7. Maneuver Alongside Subprocess

Analysis of Alternatives: Technologies and Techniques

The Navy must adopt radical approaches to find alternatives to the current UNREP process, which relies heavily on legacy equipment. This includes the need to change significant UNREP equipment aboard ships, such as the CONREP station on the flight deck, and the need for special equipment to fuel large deck ships. This research proposes substituting new automation technology for existing “analog” processes such as the phone and distance communication lines. These could be replaced by laser rangefinders that display the distance between ships on large LED displays. This would eliminate 2 to 3 personnel on the transferring ship and eliminate the need for sound-powered phones or signals. The number of personnel required for the overall personnel alongside the UNREP process can be reduced by providing more automation and analog equipment, such as wheeled transfer rails. The potential risk is that these are another thing to pack for the future alongside UNREP and have been known to break.

KVA + Integrated Risk Management Analysis Description

The KVA method aims to capture the value information within each subprocess of the UNREP process, ensuring a reliable estimate of returns on investment (ROI) and return on knowledge (ROK). The KVA analysis involves comparing inputs from subject matter experts (SMEs) to create an accurate model of the process and its subprocesses. The utilization of the KVA method in analyzing the UNREP process, specifically focusing on the As-Is baseline and To-Be optimization analyses, provides an important consideration for this research.

The ROI ratio and ROK estimate are key components of the analysis, with the ROI ratio requiring a monetized numerator and denominator for revenue-cost/cost calculation, while the ROK estimate is revenue/cost, representing the monetized benefits to cost ratio. The KVA analysis allows for comparing subprocess performance, making the baseline information comparable and facilitating a defensible assessment of potential performance improvements in the To-Be model. Common units of output, representing the value produced by a process at a specific time, are used in the KVA method and monetized using the market comparables valuation technique. This technique enables the calculation of relative total revenue for each subprocess based on revenue per unit and the number of common output units. The resulting ROI information for each subprocess is crucial in modeling the effects of subprocess optimizations across the UNREP process, aiding in forecasting potential performance enhancements.

The KVA analysis results are presented in tables for the As-Is baseline and To-Be optimization analyses, which are then used in the Integrated Risk Management (IRM) method to determine each process optimization element’s forecasted real option value. These real options results are incorporated into portfolios, considering the level of risk associated with each real option. The combination of KVA and IRM methods has been employed for nearly 2 decades to optimize DoD and naval processes. Endnotes providing basic assumptions for the KVA analyses are included, and a more detailed description of the combined KVA+IRM method can be found in the report’s appendix. All KVA spreadsheets are available upon request, emphasizing transparency and accessibility of the analysis results.

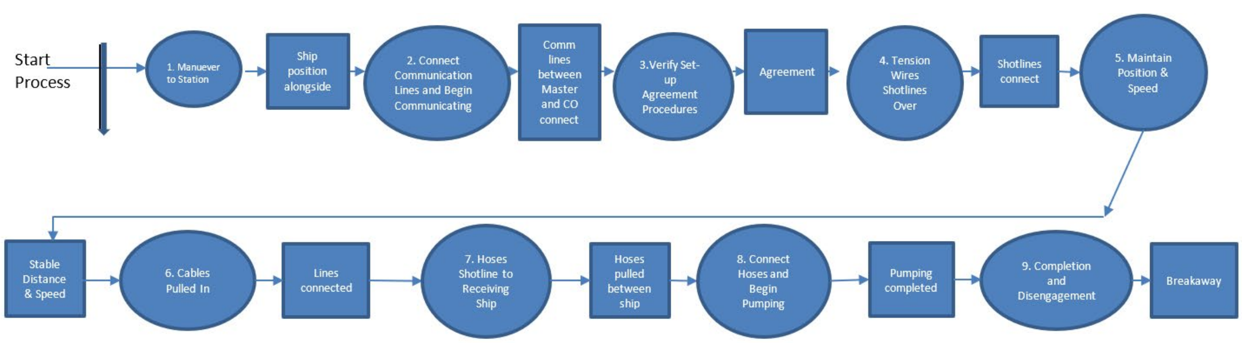


Figure 8. Along Side Process: As-Is Optimization

Table 1: Along Side UNREP KVA As-Is Analysis I

Process	TLT	Total K	Expenses	Revenues	Denominator	Numerator	ROK	ROI
1. Maneuvering to Station	26400.00	26400.00	$ 60,930	$ 242,789	$ 60,930	$ 242,789	398%	298%
2. Comm lines connected	5280.00	5280.00	$ 10,508	$ 48,558	$ 10,508	$ 48,558	462%	362%
3.Set-up Agreement Procedures	35200.00	35200.00	$ 23,654	$ 323,719	$ 23,654	$ 323,719	1369%	1269%
4.Tension lines received	17600.00	17600.00	$ 8,326	$ 161,859	$ 8,326	$ 161,859	1944%	1844%
5. Maintain Distance and Speed	35200.00	35200.00	$ 49,378	$ 323,719	$ 49,378	$ 323,719	656%	556%
6. Cable Pulled In	8800.00	8800.00	$ 9,433	$ 80,930	$ 9,433	$ 80,930	858%	758%
7. Hose Shotline Over	8800.00	8800.00	$ 26,940	$ 80,930	$ 26,940	$ 80,930	300%	200%
8. Connect Hoses	3520.00	3520.00	$ 8,478	$ 32,372	$ 8,478	$ 32,372	382%	282%
9. Completion and Disengagement (Breakaway)	32160.00	32160.00	$ 8,055	$ 295,761	$ 8,055	$ 295,761	3672%	3572%
Total	140800.00	172960.00	$197,647	$ 1,294,875	$ 197,647	$ 1,294,875	655%	555%

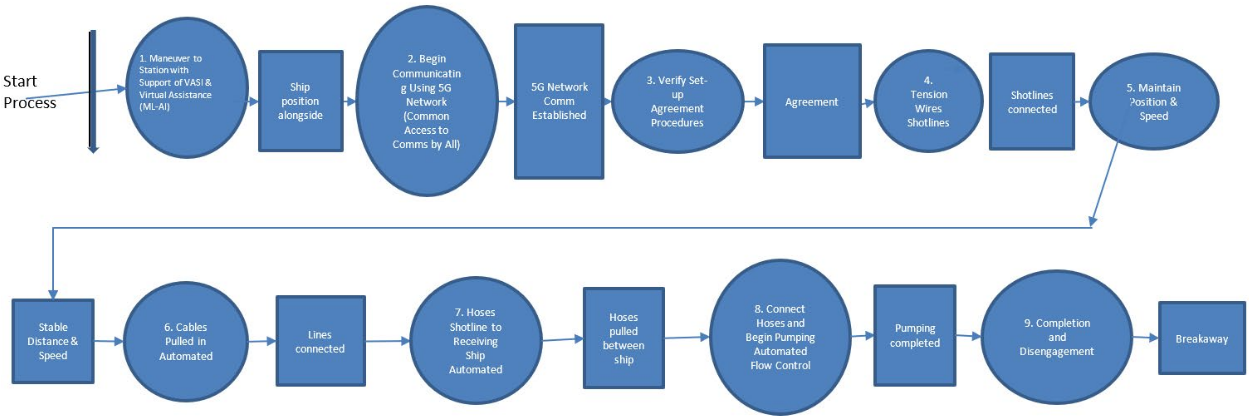


Figure 9. Along Side Process: To-Be Optimization

Table 2: Along Side Process To-Be Optimization II

Process	TLT	Total K	Expenses	Revenues	Denomiator	Numerator	ROK	ROI
1. Manuevering to Station	36000	36000	$ 16,489	$ 331,200	$ 16,489	$ 331,200	2009%	1909%
2. Comm lines connected	5280	5280	$ 8,631	$ 48,576	$ 8,631	$ 48,576	563%	463%
3.Set-up Agreement Procedures	35200	35200	$ 21,263	$ 323,840	$ 21,263	$ 323,840	1523%	1423%
4.Tension lines received	28800	28800	$ 311	$ 264,960	$ 311	$ 264,960	85236%	85136%
5. Maintain Distance and Speed	35200	35200	$ 16,459	$ 323,840	$ 16,459	$ 323,840	1968%	1868%
6. Cable Pulled In	8800	8800	$ 311	$ 80,960	$ 311	$ 80,960	26044%	25944%
7. Hose Shotline Over	8800	8800	$ 9,090	$ 80,960	$ 9,090	$ 80,960	891%	791%
8. Connect Hoses	6080	6080	$ 544	$ 55,936	$ 544	$ 55,936	10287%	10187%
9. Completion and Disengagement (Breakaway)	32160	32160	$ 5,155	$ 295,872	$ 5,155	$ 295,872	5739%	5639%
Total	164160	196320	$ 73,098	$ 1,510,272	$ 73,098	$ 1,510,272	2066%	1966%

Table 3. Assumptions and Analysis of the As-Is to To-Be Potential Optimizations

1. Maneuvering to Station: The 1611% net improvement in this subprocess is expected to be a result of implementing an automated approach system that provides an automated rudder and power setting to the receiving ship (comparable to an instrument landing system and autopilot to landing aircraft) which would increase speed of approach and decrease the need for constant bridge watch team having to monitor and correct the approach. This potential automation would increase this subprocess' information value and substantially reduce its cost.
2. Comm lines connected: There was no significant improvement, 101%, expected in this subprocess. There was a potential minor cost reduction as a result of implementing a 5G network for all watch standers.
3.Set-up Agreement Procedures: There was a minor improvement, 127%, expected in this subprocess as a result of implementing by completing this subprocess further in advance of the actual alongside UNREP. This potential optimization would provide an incremental cost reduction for this subprocess.
4.Tension lines received: Reengineering of this subprocess is expected to result in a substantial, 83292%, optimization as a result of implementing automated firing and receiving shot lines which would dramatically increase the information value and decrease the cost for this subprocess. The forecasted overall process optimization improvement was very sensitive to this potential subprocess reengineering.
5. Maintain Distance and Speed: The net improvement in this subprocess, 1312%, is expected to be a result of implementing automated helm and power settings of the receiving ship in this subprocess.
6. Tension Cable Pulled In: Reengineering of this subprocess is expected to result in a substantial, 25186% improvement in this subprocess as a result of implementing newly designed king post tensioning system either by moving king posts, i.e., in or out, or by varying king post height. This potential optimization is expected to provide a substantial cost reduction.
7. Hose Shotline for Fuel Hoses Over: The net improvement, 591%, in this subprocess is expected to be a result of implementing an automated shot line firing and recovering system. This would potentially increase the information value of this subprocess while also providing a substantial reduction in the cost.
8. Connect Hoses: Reengineering of this subprocess is expected to result in a substantial, 9905%, improvement in this subprocess as a result of implementing an automated system to extend hoses via the tension wires. This potential optimization would result in a substantial reduction in cost.
9. Completion and Disengagement (Breakaway): The potential net improvement, 2067%, in this subprocess is expected to be a result of implementing an automated low risk maneuvering capability. This optimization would substantially reduce the cost of this subprocess.

Results And Conclusions

The KVA analysis revealed a significant potential increase in ROI, primarily driven by cost optimizations in subprocesses and enhanced knowledge expertise from new automation. The analysis highlighted the value of knowledge embedded in automation and its impact on process optimization. The To-Be model's subprocess optimizations formed the basis for IRM analysis, which assessed the risks associated with each optimization option. The IRM results provided a more thorough estimate of potential values and aided sponsor leadership in investment decision-making by considering risk–reward benefits. The research proposed advanced analytical modeling processes integrating IRM and Strategic Real Options to quantify and optimize a framework for estimating and hedging prototyping uncertainty. Monte Carlo simulations were used to determine the Expected Value of Information and the value of prototyping, assisting decision-makers in understanding the criticality of prototyping. A tornado static sensitivity analysis identified critical success factors for ROI, leading to 100,000 Monte Carlo risk simulation trials comparing As-Is and To-Be models. The simulations showed a significant improvement in ROI with the To-Be model, indicating an average net ROI improvement of 843%. In most of the other subprocesses, the ROIs increased due to cost improvements.

1

The KVA analysis demonstrated the potential for a substantial increase in ROI through cost optimizations and enhanced knowledge expertise from new automation. The analysis preserved the information value of subprocesses and quantified the value of knowledge embedded in automation. The To-Be model's subprocess optimizations formed the basis for IRM analysis, which assessed risks associated with each optimization option. The IRM results provided a more thorough estimate of potential values and aided sponsor leadership in investment decision-making by considering risk–reward benefits. The research proposed advanced analytical modeling processes integrating IRM and Strategic Real Options to quantify and optimize a framework for estimating and hedging prototyping uncertainty. Monte Carlo simulations were used to determine the Expected Value of Information and the value of prototyping, assisting decision-makers in understanding the criticality of prototyping.

A tornado static sensitivity analysis identified critical success factors for ROI, leading to 100,000 Monte Carlo risk simulation trials comparing As-Is and To-Be models. The simulations showed a significant improvement in ROI with the To-Be model, indicating an average net ROI improvement of 843%. For an As-Is and To-Be 90% confidence interval, the simulated Option ROI shows a 525% and 588% ROI on As-Is and a marked improvement to 1,059% and 1,782% ROI (Figure 11). The difference is an average net ROI improvement of 843% if the To-Be improvements are incorporated (Figure 12).

This risk–reward analysis was based on inputs from the process SMEs' inputs and may be subject to updates from the given priorities of the sponsor. There may also be further, more detailed inputs to the KVA+IRM analysis based on ongoing research at the Naval Warfare Center in Point Hueneme. Further research would be required to solicit and review these various input sources for our models. However, our estimates support the potential performance improvements alongside the UNREP process when aided by potential new automation.

Future Research and Limitations

The UNREP process involves various considerations, including conducting UNREP to service the fleet in a distributed force context, ensuring sufficient LOG ships, and considering possible formations and distributions of the fleet. Improvements in naval operational efficiency are being made, including fuel, personnel, hybrid systems, and automation. Including Allied capabilities is also important, as it is expected that we will need to rely more on allies' contributions in future naval combat scenarios. This could be achieved by providing host nations with smaller but quicker turn logistics support vessels, which can be used as refueling platforms. This would require optimizing fuel flow forward and rethinking CONOPS of potential daisy-chain operations. The study did not explicitly consider other technologies required to refine the coming alongside process. A two-dimensional version of the Instrument Landing System could be put onboard receiving ships to guide them more efficiently alongside UNREP. Another potential implementation is for the replenishment ship and receiving ship to be connected in an "auto-drive" capability, which could reduce danger and improve seakeeping in higher sea states.

[1] We did not incorporate the cost of the new automation as it has yet to be selected and it would have been spread like peanut butter across the entire process making it a constant. The team at Naval Warfare Center, Point Hueneme has been working on a series of new automation capabilities for the alongside UNREP process, and when these trials are successfully completed, they estimated that the overall cost of the process would be halved (per their website information).

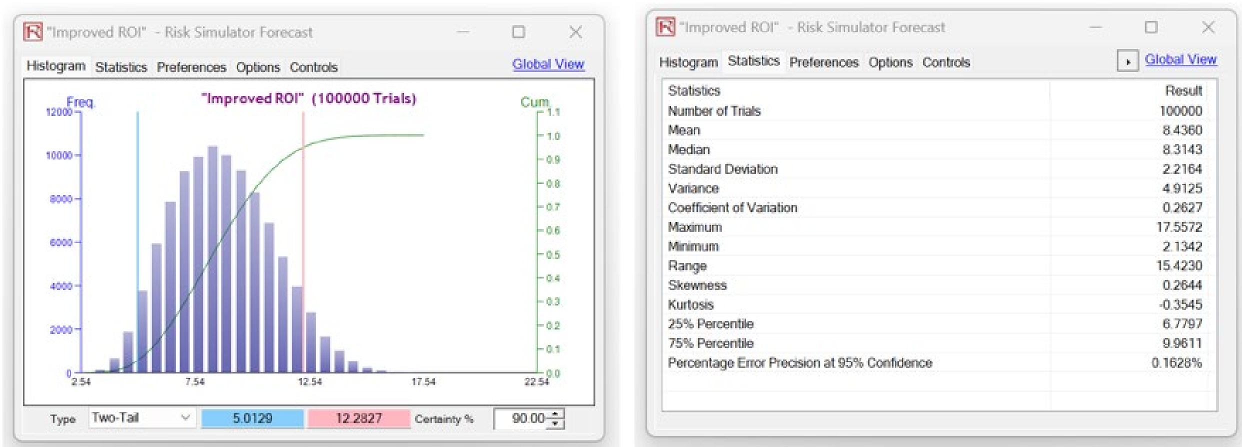


Figure 10. Simulated Option Return on Investment for As-Is and To-Be

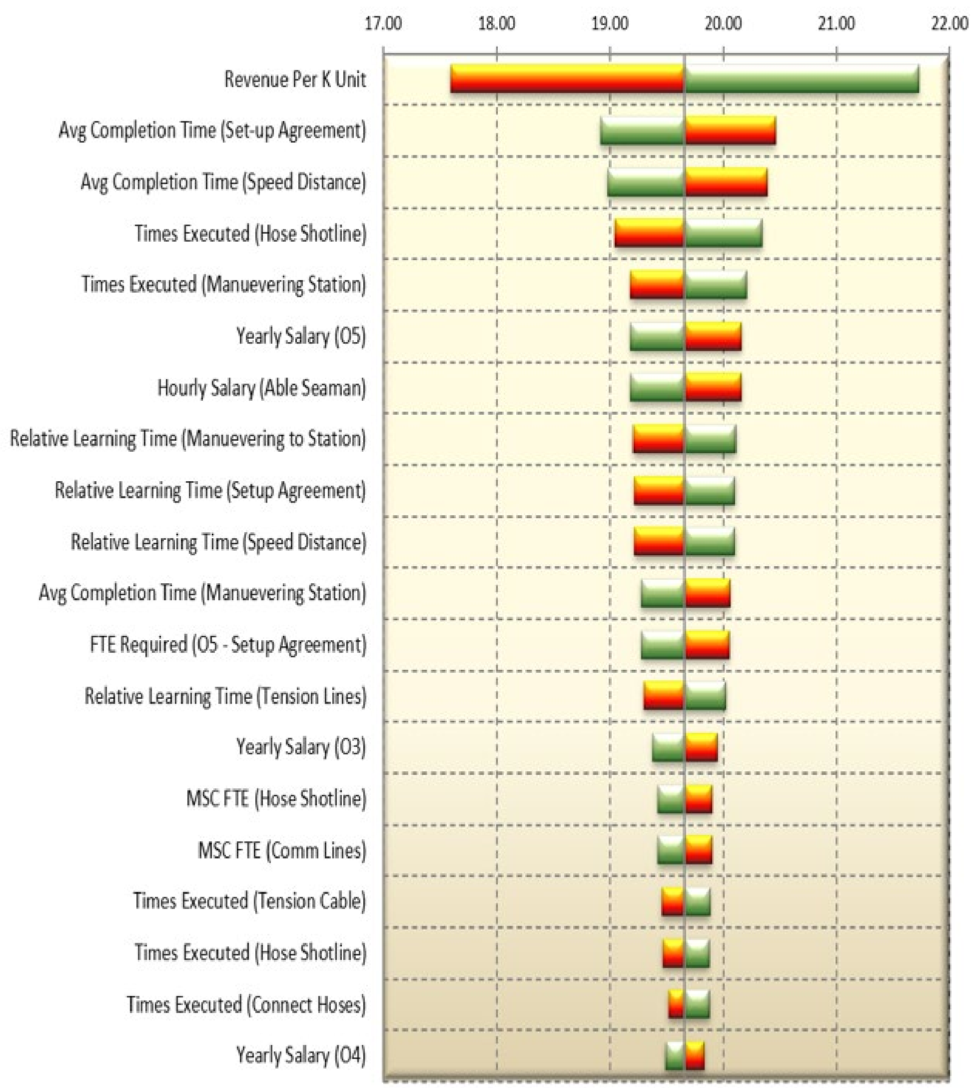


Figure 11. Tornado Static Sensitivity

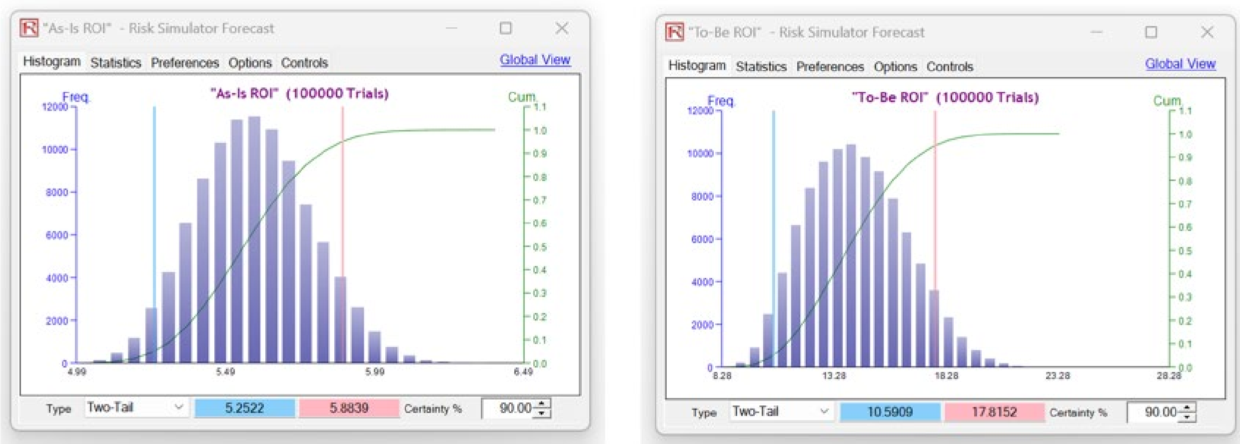


Figure 12. Simulated Option Return on Investment Increment

Appendix A: Assumptions for To-Be Calculations

- All basic assumptions about potential optimizations of the subprocesses are based on SMEs estimates and can be updated over time as the alongside subprocess new automation and procedures are implemented.
- The cost of implementing new automation for the alongside UNREP process at Port Hueneme is expected to be spread out and amortized over a 10–15-year period. The team developing the automation was unable to provide cost estimates for the current automation being tested, likely due to ongoing testing. Future studies may provide more accurate cost data for the new automation.
- The market comps estimates for the ROI surrogate revenue values were based on a 1.7 acceleration for sailor salaries in addition to a 1.5 or 50% increase of the total price that the market would pay to obtain the same process outputs.
- A day is equal to 8 working hours; 40 working hours per week; 4 weeks per month; 365 days per year.
- Rank Order of Difficulty to learn has 8 as the most difficult to learn, and 1 is the least difficult to learn.
- There are 260 working days per year.
- Automation was treated as a constant of 10% for knowledge of how to use the basic MS Office suite.
- Hourly wage is calculated from the 2023 basic pay table provided by DFAS.
- Hourly Wage is equal to the Basic Pay divided by working days times the amount of hours worked in a day.
- The revenue numerator equals the ratio of knowledge of each process to total knowledge multiplied by the total revenue.
- Times used per year are based on the number of UNREPS.
- All sailors possess a minimum of a high school reading level and several years of experience.
- All sailors possess basic office suite knowledge (Outlook/Word/Excel/PowerPoint...)
- Hourly wages based on the basic pay rate without BAH.
- Infrastructural cost, i.e., the cost of ships, fuel, hoses, etc.**. This cost was considered as a constant for the average UNREP and was not included in the cost for an average UREP. The focus was on personnel and automation costs.
- Asked Subject Matter Experts to verify or change estimates.
- The price per unit of output, K-unit, remained constant while the cost per common unit of output varied. This information will be used for future research on portfolio optimization, focusing on

potential new automation like the UNREP process improvements developed by the Port Hueneme team. Learning time estimates are a combination of formal and OJT.
- The average learner has an OOD underway, a qualified bosun's mate, and a supply officer.
- Correlation Rank order = 93%
- To learn how to do the Replenishment process (hours) = Jr officer deck qual = 200; Officer of deck qual = 300 and Actual Average LT
- Underway replenishment JOOD = 100; qualified underway OOD = 50
- The ship Captain and XO costs are not part of the cost estimate.

Appendix B: Assumptions for As-Is calculations

- The market comps included a 1.7 acceleration for salary and a 1.5 or 50% increase of the total price that the market would pay to obtain the same outputs.
- A day is equal to 8 working hours; 40 working hours per week; 4 weeks per month; 365 days per year.
- Rank Order of Difficulty to learn has 8 as the most difficult to learn, and 1 is the least difficult to learn.
- There are 260 working days per year.
- Automation is a constant of 10% for knowledge of how to use the basic MS Office suite.
- Hourly wage is calculated from the 2023 basic pay table provided by DFAS.
- Hourly Wage is equal to the Basic Pay divided by working days times the amount of hours worked in a day.
- The revenue numerator is equal to the ratio of knowledge of each process to total knowledge multiplied by the total revenue.
- Times used per year are based on the number of UNREPs.
- All sailors possess a minimum of a high school reading level.
- All sailors possess basic office suite knowledge (Outlook/Word/Excel/PowerPoint...)
- Hourly wages based on the basic pay rate without BAH.
- Infrastructural cost, i.e., the cost of ships, fuel, hoses, etc.**. This cost was considered as a constant for the average UNREP and was not included in the cost for an average UNREP. The focus was on personnel and automation costs.
- Asked Subject Matter Experts to verify or change estimates.
- The price per unit of output, K unit, was constant because all the common output units were the same. However, the cost per common unit of output varied. These assumptions allow portfolio optimization of future real options for improving the UREP process.
- Learning time estimates are a combination of formal and OJT.
- The average learner has an OOD underway, a qualified bosun's mate, and a supply officer.
- Jr officer deck qual = 200; Officer of deck qual = 300
- To learn how to do Replenishment process (hours) = Underway replenishment JOOD = 100; qualified underway OOD = 50
- The ship Captain and XO costs are not part of the cost estimate.

Appendix C: U.S. Armed Forces classes of supply

Class I – Rations – Subsistence (food and drinking water), gratuitous (free) health and comfort items.
Class II – Clothing and Equipment – individual equipment, tentage, some aerial delivery equipment, organizational tool sets and kits, hand tools, unclassified maps, administrative and housekeeping supplies, and equipment.
Class III – POL – Petroleum, Oil and Lubricants (POL) (package and bulk): Petroleum, fuels, lubricants, hydraulic and insulating oils, preservatives, liquids and gases, bulk chemical products,

coolants, deicer and antifreeze compounds, components, and additives of petroleum and chemical products, and coal.
Class IV – Construction materials, including installed equipment and all fortification and barrier materials.
Class V – Ammunition of all types, bombs, explosives, mines, fuses, detonators, pyrotechnics, missiles, rockets, propellants, and associated items.
Class VI – Personal demand items (such as health and hygiene products, soaps and toothpaste, writing material, snack food, beverages, cigarettes, batteries, alcohol, and cameras—nonmilitary sales items).
Class VII – Major end items such as launchers, tanks, mobile machine shops, some parachute systems, and vehicles.
Class VIII – Medical material (equipment and consumables), including repair parts that are particular to medical equipment. (Class VIIIa – Medical consumable supplies not including blood & blood products; Class VIIIb – Blood & blood components (whole blood, platelets, plasma, packed red cells, etc.).
Class IX – Repair parts and components to include kits, assemblies, and subassemblies (repairable or non-repairable) required for maintenance support of all equipment.
Class X – Material to support nonmilitary programs such as agriculture and economic development (not included in Classes I through IX).
Miscellaneous – Water, salvage, and captured material. Source: (AR-710-2, 2008)

References

Chen, G. R., & Fang, M. C. (2001). Hydrodynamic interactions between two ships advancing in waves. *Ocean Engineering*, *28*(8), 10532–21078. https://www.sciencedirect.com/science/article/abs/pii/S0029801800000421

Curtin, M. J. (2001). *Analysis of inter/intra ship materiel movement in sea-based logistics using simulation*. Naval Postgraduate School. https://apps.dtic.mil/sti/pdfs/ADA395806.pdf

DoN. (2001). *Underway replenishment* (NWP 4-01.4). https://maritime.org/doc/pdf/unrep-nwp04-01.pdf

Dunn, J. S. (1992). *Scheduling underway replenishment as a generalized orienteering problem*. Naval Postgraduate School. https://apps.dtic.mil/dtic/tr/fulltext/u2/a256205.pdf

Fu, S. H. S., & Haddad, W. M. (2003). Nonlinear adaptive tracking of surface vessels with exogenous disturbances. *Asian Journal of Control*, *5*(1), 88–103. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1934-6093.2003.tb00100.x

Hewgley, C., & Yakimenko, O. (2009, May). Precision-guided airdrop for vertical replenishment of naval vessels. In *20th AIAA Aerodynamic Decelerator Systems Technology Conference and Seminar* (p. 2995). https://apps.dtic.mil/sti/pdfs/ADA517932.pdf

Koteskey, T. (2020). *Spending towards strategy: Restructuring the US Navy to support offshore balancing*. https://jqas.org/wp-content/uploads/2021/01/MPF-Paper-Tyler-Koteskey.pdf

Kyrkjebø, E., Wondergem, M., Pettersen, K. Y., & Nijmeijer, H. (2004). Experimental results on synchronization control of ship rendezvous operations. *IFAC Proceedings Volumes*, *37*(10), 453–458.

Military Analysis Network. (1999, March 6). *Underway replenishment (UNREP): Navy ships*. Federation of American Scientists. https://man.fas.org/dod-101/sys/ship/unrep.htm

Miller, M. O., Hammett, J. W., & Murphy, T. P. (1987). The development of the US Navy underway replenishment fleet. *Society of Naval Architects and Marine Engineers-Transactions*, *95*. https://trid.trb.org/view/403682

Rowling, S. J., Teo, M., & West, M. (2019). *Multi-stock replenishment of naval vessels by specialist replenishment ships*. https://www.doi.org/10.36334/modsim.2019.B2.rowling

United States Department of the Army. (2008). *Inventory management: Supply policy below the national level* (Army Regulation 710-2).

https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/r710_2.pdf

PANEL 18. PLANNING, PROGRAMMING, BUDGETING, AND EXECUTION (PPBE) REFORM COMMISSIONERS' REPORT

Thursday, May 9, 2024	
12:45 p.m. – 2:00 p.m.	**Chair: Dov S. Zakheim,** Former Under Secretary of Defense (Comptroller) and Chief Financial Officer, Department of Defense **Panelists:** **Hon. Jamie Morin, Ph.D.,** Vice President, Defense Systems Operations in the Defense Systems Group, The Aerospace Corporation **John Terence Blake, VADM, USN (Ret.),** Professor of the Practice, Naval Postgraduate School, former Deputy Chief of Naval Operations for Integration of Capabilities and Resources (OPNAV N8) **Ms. Lara Sayer,** Executive Director for the Commission on Planning, Programming, Budgeting and Execution (PPBE) Reform **Ms. Elizabeth Bieri,** Director of Research for the Commission on Planning, Programming, Budgeting and Execution (PPBE) Reform

Dov S. Zakheim—is Senior Advisor at the Center for Strategic and International Studies and Senior Fellow at the CNA Corporation, a federally funded think tank. Previously he was Senior Vice President of Booz Allen Hamilton where he led the Firm's support of U.S. Combatant Commanders worldwide.

From 2001 to 2004 he was Under Secretary of Defense (Comptroller) and Chief Financial Officer for the Department of Defense, and from 2002-2004 he was also DOD's coordinator of civilian programs in Afghanistan. From 1985 until 1987, Dr. Zakheim was Deputy Under Secretary of Defense for Planning and Resources. He held other senior DOD posts from 1981-1985.

Hon. Jamie Morin, Ph.D—Jamie M. Morin is vice president of Defense Strategic Space at The Aerospace Corporation. As part of Aerospace's Defense Systems Group, Jamie leads technical support to the seniormost levels of the Department of Defense and Department of the Air Force, including the U.S. Space Force and U.S. Air Force headquarters, as well as to the other military services and combatant commands.

Morin also is executive director of the Center for Space Policy and Strategy, which provides objective analysis to ensure well-informed, technically defensible, and forward-looking space policy across the civil, military, intelligence, and commercial space sectors. In that role, he has published research in outlets including Nature, Defense One, Defense News, and the Journal of Space Safety Engineering. He speaks widely and orchestrates the Center's extensive series of publications, events, and multimedia products to shape the future of the U.S. space enterprise.

From 2017 to 2023, Morin was vice president of Defense Systems Operations at Aerospace.

Prior to joining Aerospace, Morin served as director of Cost Assessment and Program Evaluation (CAPE) for the Department of Defense, where he led the organization responsible for analyzing and evaluating the department's plans, programs, and budgets in relation to U.S. defense objectives, threats, estimated costs, and resource constraints.

Before his appointment as director of CAPE, Morin served for five years as the assistant secretary of the Air Force (Financial Management and Comptroller). He also served for a year as acting under secretary of the Air Force, where he led the Air Force Space Board and the Air Force Council.

Before moving to the Pentagon, Morin was lead analyst for defense, intelligence, and foreign affairs on the professional staff of the U.S. Senate Committee on the Budget.

In addition to his role at Aerospace, Morin is an adjunct professor of international relations at Georgetown University. He also serves in various advisory roles to the government, including the Secretary of State's International Security Advisory Board, the Secretary of the Air Force's senior advisory group, and the Commission on the Reform of the Planning, Programming, Budgeting and Execution System.

John Terence Blake, VADM, USN (Ret.)—was appointed to the United States Naval Academy from the state of New York, he graduated in 1975. His sea duty assignments include: USS New (DD 818), USS Sarfield (DD 837), USS Joseph Strauss (DDG 16), USS John Young (DD 973), USS Chandler (DDG 996), USS Leahy (CG 16), and USS Blue Ridge (LCC 19).

Blake commanded the destroyer USS O'Brien (DD 975), served on the 7th Fleet Staff as current operations and assistant chief of staff for Operations, commanded the guided-missile cruiser USS Normandy (CG 60) and served as commander, Carrier Strike Group 11.

His shore duty assignments include: flag lieutenant to commander, Navy Recruiting Command; Naval Post Graduate School where he earned a masters degree in Finance; Navy Staff (N80) head, Sea Control Section and program manager for the Navy Shipbuilding account; National War College where he earned a masters degree in National Security; Joint Staff (J8) division chief and head of the Combat Identification Joint Warfare Capability Assessment Team; director, Programming Division (N80); director, Operations Division, Office of Budget in the Office of the Assistant Secretary of the Navy (Financial Management/Comptroller); director, Operations Division, Fiscal Management Division in the Office of the Chief of Naval Operations; deputy director for Resources and Acquisition on the Joint Chiefs of Staff (J8) and deputy assistant secretary of the Navy for Budget.

Blake was assigned as deputy chief of Naval Operations, Integration of Capabilities and Resources in Washington. He retired in February 2013.

He is authorized to wear the Navy Distinguished Service Medal, Defense Superior Service Medal with oak leaf cluster, the Legion of Merit with four gold stars, the Meritorious Service Medal with two gold stars, the Navy and Marine Corps Commendation Medal with two gold stars and various service and campaign medals.

Lara Sayer—is the Executive Director for the Commission on Planning, Programming, Budgeting and Execution (PPBE) Reform. Prior to joining the Commission, Ms. Sayer was a member of the Senior Executive Service and served as the Comptroller for Commander, Navy Installations Command (CNIC) and Naval Facilities Engineering Systems Command (NAVFAC. Ms. Sayer also served as the Comptroller for the Office of Naval Research (ONR) and the Deputy Chief Financial Officer for United States Special Operations Command. Ms. Sayer has also served in several senior positions within the Air Force in the acquisition, budget, and resource management sectors. Ms. Sayer holds a Bachelor of Music in Vocal Performance and a Master of Business Administration from Wright State University. She also has a Masters in National Resource Strategy from the Eisenhower School at the National Defense University.

Ms. Elizabeth Bieri—is the Director of Research for the Commission on Planning, Programming, Budgeting and Execution (PPBE) Reform. Ms. Bieri has spent her career in financial management, logistics, and as a cost and price analyst in the national defense space having served in the Army G-4, Assistant Secretary of the Army for Acquisition, Logistics and Technology, and other commands within the Army Materiel Command enterprise; the Marine Corps Deputy Commandant for Programs and Resources; and the Office of the Under Secretary of Defense for Comptroller. Prior to this position, she was the Deputy Director for Budget in the Office of the Assistant Secretary of Defense for Special Operations and Low-Intensity Conflict, Secretariat for Special Operations. Ms. Bieri holds a Master of Arts in National Security and Strategic Studies from the U.S. Naval War College and a Bachelor of Arts from Grinnell College.

PANEL 19. ADVANCEMENTS IN DIGITAL ENGINEERING FOR TEST AND EVALUATION

Thursday, May 9, 2024	
2:15 p.m. – 3:30 p.m.	**Chair: Christopher C. Collins,** Executive Director, Developmental Test, Evaluation, and Assessments ***Model-Based Integrated Decision Support Key: A Standardized Approach to Mitigating Decision Support Challenges During Acquisition Test and Evaluation*** Awele Anyanhun, GA Tech ***Accelerating Implementation of Critical Joint Warfighting Concepts and Capabilities*** Craig Arndt, GA Tech Jeremy Warner, DoD DOT & E ***Advantages of Using Complex Decision Support Tools in Planning Multi-Modal Test Programs*** Milo Taylor, GA Tech

Christopher C. Collins—is the Executive Director, Developmental Test, Evaluation, and Assessments (ED,DTE&A) within the Office of the Undersecretary of Defense Research and Engineering (OUSD(R&E)). DTE&A supports the Department of Defense's acquisition programs using innovative and efficient DT&E strategies to ensure production readiness and fielded systems to meet Warfighter/User needs; improve the Defense Acquisition T&E workforce "practice of the profession;" and advance T&E policy and guidance. DTE&A also conducts Independent Technical Review Assessments (ITRA) and Milestone Assessments for major acquisition programs.

Mr. Collins was appointed to the Senior Executive Service in April 2020. Prior to his appointment, he was the COMNAVSEASYSCOM Deputy for Test and Evaluation. He has also served in various engineering and test leadership positions in the Aegis Ballistic Missile Defense Program within the Missile Defense Agency. He also completed a one-year experiential assignment with the United States Air Force on the Headquarters Staff.

Mr. Collins career began with a United States Naval Academy commission in 1984. He completed a combined Active Component and Reserve Component career and retired after 30 years at the rank of Navy Captain. While on active duty, he completed several deployments as a Navy helicopter pilot. He supported Navy technology transition initiatives and assessments at the Office of Naval Research while on reserve duty. During his reserve tenure, he held command of two Reserve Component Commands.

Mr. Collins has a Bachelor of Science in Electrical Engineering from the United States Naval Academy and a Master of Science in Aeronautical Engineering from the Naval Postgraduate School. He graduated with distinction from the Navy Command and Staff College (distance education) and the Air War College (in-resident). He is a graduate of the 2016 cohort of the Defense Senior Leader Development Program. He is a member of the Defense Acquisition Corps. He has achieved Level III Certification in Program Management, Engineering, and Test and Evaluation.

Model-Based Integrated Decision Support Key: A Standardized Approach to Mitigating Decision Support Challenges During Acquisition Test and Evaluation

Dr. Awele Anyanhun— is a Senior Research Engineer in the Enterprise and Open Architecture (EOSA) Branch at Georgia Tech Research Institute (GTRI). She is an INCOSE-Certified Systems Engineering Professional (CSEP) and Senior Member of IEEE with over 13 years' professional experience in architecting complex automotive, space, and defense system architectures. Anyanhun is an OMG-Certified Systems and Software Modeling Professional (SysML-MBA, OCUP 2-MBA) and a UL-Certified Functional Safety Professional (UL-CFSP). Anyanhun has authored multiple conference and journal publications, and holds a PhD in Electrical Engineering with a concentration in model-based systems engineering.. [Awele.Anyanhun@gtri.gatech.edu]

Dr. Craig Arndt— has extensive experience as a senior executive and technology leader in research, education, engineering and defense, homeland security and intelligence technologies, with extensive experience as an innovative leader in industry, academia, and government. Arndt currently serves as a Principal Research Engineer on the faculty of the George Tech Research Institute (GTRI) in the System Engineering Research Division of the Electronic Systems Lab. Arndt is a licensed Professional Engineer (PE), a Certified Human Factors Professional (CHFP), and an Expert Systems Engineering Professional (ESEP). He has over 40 years of professional engineering experience though the defense and government engineering community. He is widely published in the areas of electrical, systems, and human factors engineering and serves on the boards of several technical organizations. Arndt holds engineering degrees in electrical engineering, systems engineering, and human factors engineering and a Master of Arts in strategic studies from the U.S. Naval War College. [Craig.Arndt@gtri.gatech.edu]

Abstract

Providing timely decision support to decision-making authorities during the various phases of an acquisition program is critical for the on-time delivery of operationally effective weapon systems that meet the needs of the warfighter. To ensure decision-makers are equipped with the necessary test and evaluation (T&E) data to inform decisions, the Department of Defense (DoD) recently mandated the use of the Integrated Decision Support Key (IDSK) as a tool to encapsulate (i.e., succinctly record) a program's decisions and the T&E data necessary to support the decisions. Therefore, an approach that utilizes digital engineering, specifically model-based systems engineering (MBSE) as a means to standardize the linkage of test data to decisions presents a significant value proposition for decision-making authorities—linking data from a program's system, design, and test planning models to key acquisition decisions. An overt value of this approach is the resulting digital thread that connects data sources (i.e., digital models) into an authoritative source of truth to both inform and validate decisions. Hence, this paper presents a Model-Based Integrated Decision Support Key (MB-IDSK) Reference Architecture (RA) that integrates and links data from multiple digital models to a standardized set of acquisition, technical, and T&E decisions. The MB-IDSK RA provides a standardized pattern and approach for developing program-specific MB-IDSKs to support program acquisition and T&E decision-making.

Keywords: Acquisition Decision Support, IDSK, Model-Based Systems Engineering, Reference Architecture

Introduction

Department of Defense (DoD) decision-making authorities across acquisition programs are expected to make decisions that are consistent, coherent, and timely to build and maintain enduring advantage in the delivery of weapon systems to the warfighter. To support timely decision-making, program and mission-critical vulnerabilities involving test planning, test prioritization, and the testing capabilities of test facilities and ranges must be identified and mitigated prior to key decision points. In order to accelerate the delivery of systems that work, it is necessary to create tools and processes that optimize integrated T&E and support the proliferation of information to decision-makers as early as possible in the acquisition lifecycle. Moreover, within the context of shrinking error margins, shorter decision-cycle times, and in the face of a growing attack-surface, providing decision support in the form of accurate and trusted data at the speed of need becomes critical.

To better support decision-making across a program's lifecycle, the traditional Integrated Decision Support Key (IDSK) was developed as a framework to identify and specify critical T&E data required to inform defense acquisition program decisions. In addition, it specifies relevant information about a program's decision-making process throughout the acquisition cycle to support decision-makers as stated in the DoD Instruction 5000.89 document (Executive Services Directorate, 2020). As a consequence of this directive, the Director, Operational Test and Evaluation (DOT&E) outlined a key strategy—accelerate the development of solutions that enable digital representations of numerous T&E tools and artifacts including a digital IDSK (Guertin, 2022). This strategy underscores a critical need which this work seeks to address by developing a digital engineering artifact in the form of a Model-Based IDSK (MB-IDSK) Reference Architecture (RA) that, when instantiated, will seamlessly integrate into the digital engineering ecosystem. The MB-IDSK RA proposed in this work provides consistency, integrity, balance, and practical guidelines for program-specific implementations. Specifically, an MB-IDSK will improve the decision-making process by making it compatible and interactive with the systems engineering models for the system under development (SUD). Additionally, a library of standardized tailorable IDSK table templates that are fully consistent with the traditional paper and table-based IDSKs used in other programs within the DoD are generated to support test planning and decision-making. The rest of the paper is organized as follows: A brief background on the IDSK is presented in Section II, while the value proposition for an MB-IDSK RA is outlined in Section III. The proposed MB-IDSK RA is described in Section IV, while an overview of how the MB-IDSK RA can be instantiated by an acquisition program is summarized in Section V. Conclusion and Future Work are presented in Section VI and Section VII.

Background

A number of research studies have been conducted on best approaches to support acquisition T&E decision-making. Beers et al. (2013) reported on the developmental test evaluation framework which describes a logical thought process involving defining an evaluation framework, building analytically test programs to generate data, and evaluating data in order to inform decisions. Also reported by Beers (2022) was the use of a digital IDSK, which focuses on gathering data to evaluate operational and technical capabilities in order to inform acquisition and operational fielding decisions. Collins and Beers (2021) explored the concept of applying the IDSK during the post-mission engineering phase in order to evaluate capabilities and inform operational fielding decision-making. Additionally, Werner and Arndt (2023) reported on the development of digital engineering artifacts to support decision-making. In more recent development, DOT&E defined a Baseline IDSK for use by acquisition programs. The Baseline IDSK comprises a series of tables in the form of *IDSK-long and IDSK-short tables, Dictionaries*, *Resource tables*, which can be implemented using a range of different technologies based on the purview of the program office (PO) and vendors involved in executing the program.

Although most research studies examined involve various approaches for improving acquisition T&E decision-making, none adequately addressed the standardization of these decisions in a repeatable consistent manner and the linking of decisions to data resident in program digital models. To the best of our knowledge, there is currently no published research work that exploits Model-Based Systems Engineering (MBSE) methods for IDSK development.

We address this gap by aligning our model-based approach with best practices from within the DoD and the systems engineering and modeling community to provide data-driven decision support using MBSE and systems modeling language (SysML).

The Value Proposition of an IDSK RA for DoD Acquisition Programs

The motivation behind defining an RA for the IDSK is based on the premise that an architecture should reflect the organization of the owning enterprise (Army Aviation and Missile Command Fort Eustis, VA [AMCOM], 2022). Therefore, for a hierarchical organization such as the DoD/DOT&E enterprise, developing an IDSK RA presents a critical first step towards preventing conflicting business objectives for programs of record (PoR) by serving as a medium to flow down the overarching business objectives for a PoR IDSK as perceived by the DoD/DOT&E authorities. Specifically, the IDSK RA represents an essential tool to facilitate communication and alignment efforts of current and future IDSK architectures. Figure 1 depicts the IDSK architecture strategy as adapted from the DoD Comprehensive Architecture Strategy.

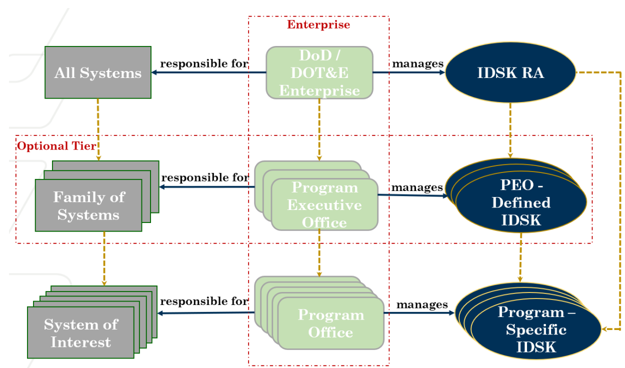


Figure 1. IDSK RA Architecture Strategy (AMCOM, 2022)

Equipping DoD acquisition programs with overarching guidance on how to leverage digital engineering for decision support is critical to achieving the enterprise-wide business and mission objectives of providing weapon systems at the speed of need and relevancy. An RA provides a method for focusing all architecture and design decisions with the intent to enforce common applicable standards and providing a tailorable architectural structure (AMCOM, 2022; Muller & Hole, 2007). The IDSK RA is developed to demonstrate and provide guidance on how the T&E enterprise and acquisition programs implementing digital engineering could leverage existing digital models created during the various acquisition phases as real-time data sources to inform key program decisions and improve decision outcomes. Figure 2 describes the role of the IDSK RA relative to program-specific IDSKs.

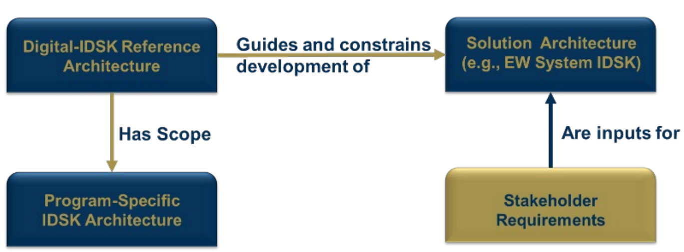


Figure 2. The Role of the IDSK RA to Program-Specific IDSKs

The three crucial characteristics that underpin the IDSK RA approach include

1) The creation of a digital thread that links acquisition and test data resident in missions, systems, and test models to metrics and key decisions.

2) Key decisions and decision classes that are standardized across acquisition T&E programs which help define expectations, formalize processes, and create accountability for programs.

3) A library of tailorable IDSK table template types—highlighted in Figure 3—and model navigation syntax in the form of query elements that are easily modified based on a program's specific implementation of the MB-IDSK RA.

IDSK RA Standardized Table Formats		
	Table Type	Count
1	IDSK RA Dictionary Standardized Table	Ten (10)
2	IDSK RA Test Resources Standardized Table	Eight (8)
3	IDSK Crosswalk Standardized Table	Eleven (11)
4	IDSK RA Key Decision Standardized Table	Five (5)

Figure 3. IDSK RA Standardized Table Types and Count

The IDSK RA

The MB-IDSK RA captures the essence of the decision support domain relative to the needs of acquisition T&E decision-makers. Specifically, it represents an instantiable pattern developed using MBSE principles and best practices to provide guidance for the development of new and/or extended versions of program-specific MB-IDSKs. In this section we describe briefly the key business and architecture drivers of the MB-IDSK RA, multiple architecture views, and a set of standardized IDSK tables generated from instantiated notional IDSK architecture exemplars.

Identifying the IDSK RA Key Business and Architecture Drivers

The intent of Key Business Drivers (KBDs) is to convey stakeholder vision, guidance, and critical business concerns; they answer "why" the architecture is needed (AMCOM, 2022). Two MB-IDSK RA KBDs—*link to digital models* and *lightweight architecture*—were determined by analyzing and prioritizing stakeholder concerns, primarily those aligned with the second pillar—*accelerate the delivery of weapons that work*—of the DOT&E strategy (Guertin, 2022). Notably, the need to leverage existing digital models (i.e., the resources of a program's digital engineering ecosystem) as data sources to provide timely decision support is a critical business concern of both acquisition programs and the T&E enterprise. Mapping of KBDs to stakeholder concerns are shown in Figure 4.

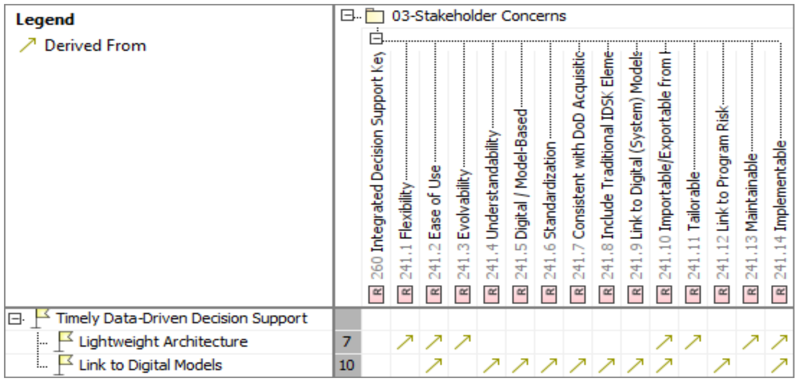


Figure 4. **Derivation of KBDs From Stakeholder Concerns**

Portrayed in Figure 5 is a tiered layout of selected MB-IDSK RA KBDs and key architectural drivers (KADs). KADs—usually expressed as architectural requirements—are a combination of business, operational (functional), quality attributes (nonfunctional) requirements, and constraints which are critical to the success of a given entity. In the case of the IDSK RA, top-level KADs—*data integration, standardization*, and *flexibility*—are vital to ensuring successful deployment (i.e., its acceptance and use by programs). Consequently, the MB-IDSK RA is able to maintain a flexible posture by fostering evolvability and responsivity and maintainability quality attributes.

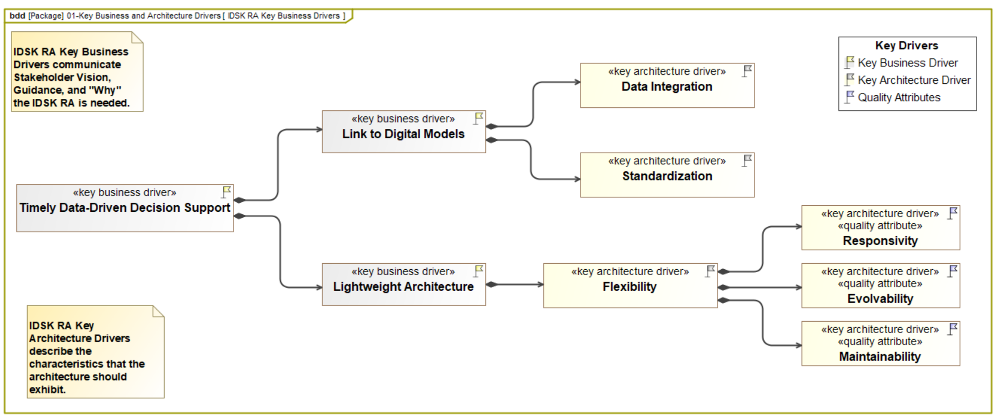


Figure 5. IDSK RA KBDs and KADs View

A. Enabling Data-Driven Decisions: Standardized IDSK Table Views

The set of standardized IDSK table formats generated from the MB-IDSK RA are the primary decision support artifacts of the IDSK RA. These table are generated from the various views specified in the IDSK RA model and collectively represent the integration of information, knowledge, capabilities, and data necessary to support decision-making by POs and the T&E enterprise to achieve their strategic objectives. A major benefit of this model-based architectural approach to decision support is the latitude it affords in generating views that can be tailored and configured based on the needs of the decision-making authority. Notional examples of several IDSK table formats are presented in this subsection. Currently, a total of 26 acquisition test planning and decision-related table views make up the model-based IDSK Tables Library.

IDSK Dictionary Standardized Format. IDSK dictionary tables defined in the IDSK RA capture information regarding key IDSK elements and their corresponding descriptions as portrayed in Table 1. Dictionary table views can be tailored to highlight additional data fields as shown in Table 2.

Table 1. Metric Dictionary Table (Notional)

#	Name	Metric Type	Metric Description
1	reliability	EW System Reliability	Metric Description/documentation goes here.
2	suitability	EW System Suitability	Metric Description/documentation goes here.

Table 2. Decision Dictionary Table (Notional)

#	Name	Decision Type	Decision Question	Decision Category
1	ctr	EW SUT Critical Performance Indicator Decision	Decision question goes here	*Critical Technical Performance Decision*
2	ctr	EW SUT Critical Performance Indicator Decision	Decision question goes here	*Critical Technical Performance Decision*
3	mpc	EW SUT Operational Availability Indicator	WILL THE SYSTEM MEET OPERATIONAL AVAILABILITY NEEDS?	*Programmatic Decision*

IDSK Test Resource Standardized Format. IDSK test resource tables capture important test planning data required to support acquisition T&E planning and decision-making. Notional examples of the IDSK RA *Test Event Resource* and *Test Article Resource* tables are depicted in Table 3 and Table 4.

Table 3. Test Event Resource Table (Notional)

#	Name	Test Event Quarter & Fiscal Year	Type of Test	Test Number	Test Objective	Test Range
1	Missile SUT Test Event	Q1FY24	Developmental Test (DT)	4.3.2	Determine if the system requirement can be met by the current design.	Missile Test Range Facility

Table 4. Test Article Resource Table (Notional)

#	Name	Test Article (SUT)	Quantity (SUT)	Support System(s) Required for Test	Quantity (Support System)	Duration_Hour	Dollar Cost	Dollar Cost Total
1	EW SUT Test Event	EW SUT Alpha	5	Test Aircraft	2	10	3000	30000
2	Missile SUT Test Event	Missile SUT	2	Nuclear Submarine	1	5	20000	65000

IDSK Crosswalk Standardized Format. IDSK crosswalk tables capture cross-cutting views which expose important dependencies between key IDSK elements, giving a holistic view of T&E data to support timely decision-making. Table 5 and Table 6 depict examples of *Decisions Crosswalk* and *Metrics Crosswalk* tables.

Table 5. Decisions Crosswalk Table (Notional)

#	Name	Metrics	Operational Requirement	Data Element	Threshold	Objective	Test Data Collected	Test Data Quantity Required	Data an Analysis URL	△ Current Estimate
1	EW SUT Sensor Requirement Decision	EW System Suitability	229.2 EW SUT Compute Target ID	TestResult1 : VerdictKind = inconclusive dataElement3 : Real = 300.0 dataElement4 : Real = 150.0	42.0 55.0	45.9 56.0	34 22	40 22	www.idsk.com www.idsk.org	4.5 6.7

Table 6. Operational Metrics Crosswalk Table (Notional)

#	△ Name	Operational Requirement	Operational Requirement Type	Derived Technical Requirement	Key Decision	Tests
1	Missile System Suitability	229.1 Missile Speed Requirement	KPP	219 Missile Speed Requirement	Missile System Critical Performance Indicator Missile System Functional Review	All Scenarios Test Missile Speed Test Scenario

IDSK Key Decision Standardized Format. The IDSK RA comprises four types of decision tables to support decision-makers throughout the acquisition and T&E process. Specifically, decision tables include *Class I—critical technical requirement*, *Class II—milestone review*, *Class III—subsystem critical performance and tech maturity*, *Class IV—operational performance characteristics*, and *Class V—programmatic decisions* tables. Shown in Table 7, Table 8, and Table 9 are the IDSK Class I, II, and III decision-type tables.

Table 7. Class I Decision Table—Critical Technical Requirement Decision Table (Notional)

#	Name	△ Program Decision	Decision Type	Decision Category	Decision Outcome	Decision Date	Milestone Review Gate	Confidence Level Required for Decision	Applicable Requirement	System Requirement	Data Required	Data Source
1	Missile System Critical Performance Indicator	CAN THIS REQUIREMENT BE MET WITH THE CURRENT SYSTEM DESIGN?	Critical Technical Requirement Decision	CLASS I	Meets Requirement	12/1/23	Preliminary Design Review	High Confidence (> 90%)	229.1 Missile Speed Requirement	219 Missile Speed Requirement	Missile Speed Test Data	Missile Speed Test

Table 8. Class II Decision Table—Milestone/Technical Review Decision Table (Notional)

#	△ Name	Decision Question	Decision Outcome	Decision Category	Decision Type	Confidence Level Required for Decision	Milestone Gate	Decision Date	KSA, KPP, KPI	System Requirement	Data Source	Data Required
1	Missile System Functional Review	IS THE MISSILE ABLE TO MAINTAIN MINIMUM SPEED BASED ON ITS FUNCTIONAL DESIGN?	Inconclusive	CLASS II	Milestone B - Engineering Development	High Confidence (> 90%)	Based on Review Schedule	11/6/24	229.1 Missile Speed Requirement	219 Missile Speed Requirement	Missile Speed Test	Missile Speed Test Functional Data

Table 9. Class III Decision Table—Subsystem Critical Performance Decision Table (Notional)

#	Name	Program Decision (Class III)	Decision Category	Decision Type	Confidence Level Required for Decision	Milestone Gate	Decision Date	Decision Outcome	Technical Measure	System Requirement	Data Source	Test Data Required
1	EW SUT Sensor Requirement Decision	WILL THE EW SYSTEM SENSORS, SIGNAL PROCESSING, AND TARGET DETECTION ALGORITHMS PERFORM UP TO THE REQUIREMENTS?	CLASS III	Sensor Requirement	High Confidence (> 90%)	Based on Review Schedule	9/19/24	Level 6 or Higher	229.2 EW SUT Compute Target ID	184 EW SUT Compute Correct Target ID	Test	Sensor Test Data

Enabling Data-Driven Decisions: IDSK RA Views and Viewpoints

The perspectives of acquisition and T&E decision-makers—*IDSK stakeholders*—form the basis for the IDSK RA viewpoints and corresponding views. A viewpoint as stated in the *Software, Systems, and Enterprise—Architecture Description ISO Standard* (ISO/IEC/IEEE, 2022) establishes the conventions for creating, interpreting, presenting, and analyzing a view to address the concerns framed by a viewpoint. The IDSK-RA description is illustrated through views depicted as diagrams. These views are created to serve as digestible chunks of the complete architecture and address specific concerns of acquisition test-planning stakeholders and decision-makers as it relates to their decision support needs.

Importantly, the IDSK-RA is developed to facilitate both current and future program-specific IDSK implementations by utilizing architecting principles such as the separation of concerns, managing key interfaces, and ensuring minimal coupling between elements. Abstractions and simplification concepts are also utilized in relation to how diagram views appear and how they are presented in this work.

Defining the IDSK RA T&E Decision Support Overarching View. An overarching view of the IDSK RA is shown in Figure 6. Although several elements and relationships have been deliberately elided from the view to enhance readability, the view still provides crucial insights into the top-level composition of the acquisition T&E decision support domain. The RA links traditional elements of the IDSK—Decisions, Data, and Data Sources (e.g., Tests)—to cardinal decision-enabling elements captured within a program's digital engineering ecosystem. Some of these elements include *metrics, test personnel, decision-makers, program office artifacts, test budget, program risk,* and data captured in *requirements, system*, and *test range* models.

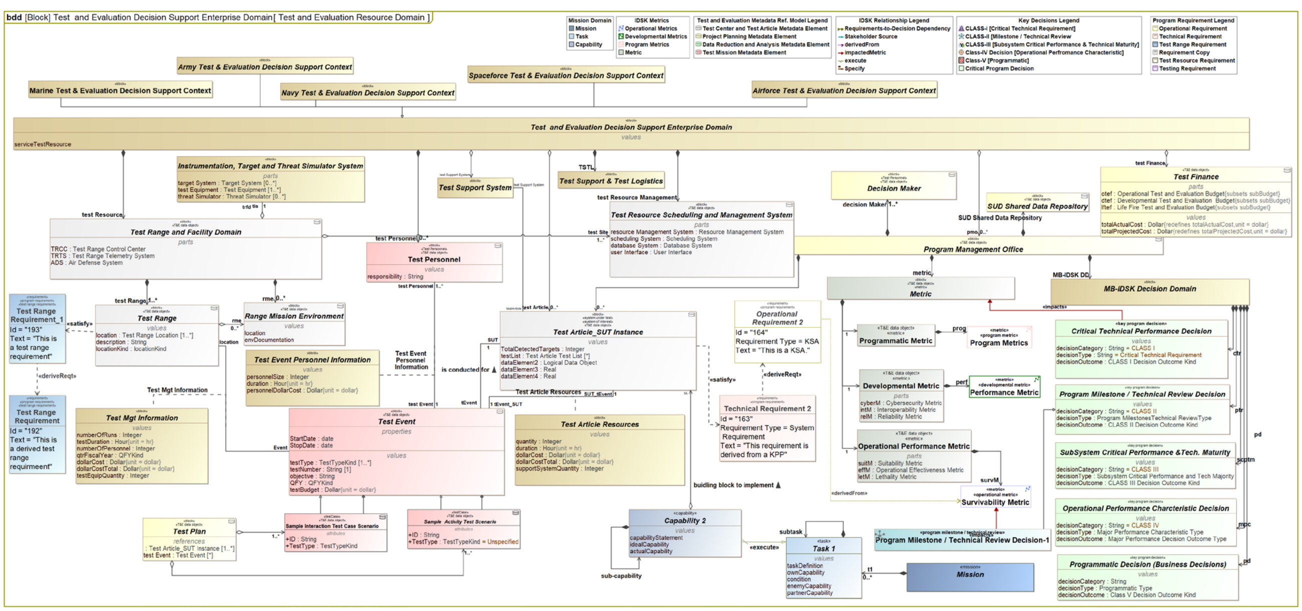


Figure 6. IDSK RA T&E Decision Support Domain View (Partial)

Key Decisions Domain Viewpoint and View. Within the context of decision-making in DoD acquisition programs, there are a limited number of critical decisions that need to be made at different times and based on different aspects of the program. These different decisions are well documented in the DoD 5000 and several other DoD acquisition process documents. To consistently make the best decisions, the availability of decisions and decision classes that are standardized across acquisition programs is necessary to help define expectations, formalize processes, and create accountability for programs.

Presently, a standardized set of program decisions grouped into five classes are defined in the IDSK RA. These decision classes provide a structured context for specifying the limited number of critical decisions that need to be made throughout the acquisition process and provide a format to link them to developmental, operational, and integrated test data needed to inform decisions. These classes are Class I, Critical Technical Requirements Decisions; Class II, Program Milestones/Technical Reviews Decisions; Class III, Sub-System Critical Performance and Technology Maturity Decisions; Class IV, Major Performance Characteristics Decisions; and Class V, Programmatic Decisions.

Figure 7 depicts the various IDSK RA categories of decisions (Class I–V) and the specific data characteristics of each decision class. Some characteristics defined as attributes include *decision question, decision outcome, confidence-level required, data source*, the specific *data required* to inform the decision, the *decision type*, and the *date* by which the decision is required amongst others. Sample instantiations of each decision class are also highlighted. The Key Decision Domain Viewpoint addresses the concern—What types of decision classes and corresponding metadata are required to support the generation of the IDSK key decision tables?

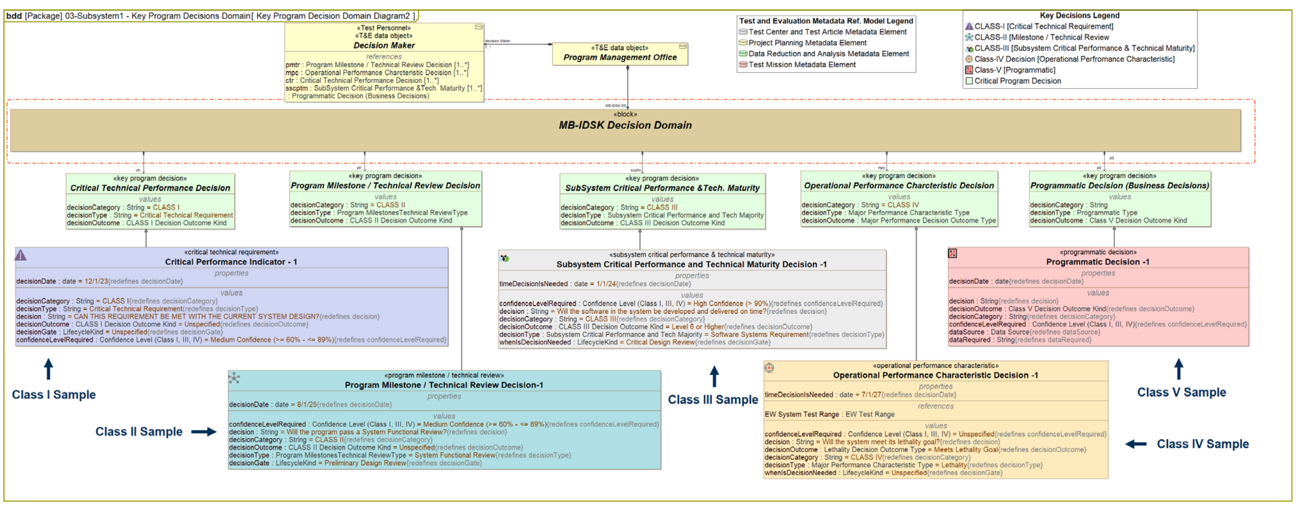


Figure 7. IDSK RA Decisions Domain View

Metrics Domain Viewpoint and View. The Metrics Viewpoint defines the IDSK RA Metrics-types required for evaluating the system under test (SUT) during the various phases of system development and test. These metrics are crucial to assisting decision-makers make the best decisions. Figure 8 highlights a Metrics View of the IDSK RA and portrays the key relationships between the Metrics and other key elements of the IDSK RA, which include the operational requirements—derived from the metrics—and the critical program decisions which impact the metrics. Three main classes of metrics currently specified the IDSK RA include *operational metrics*, *developmental metrics*, and *programmatic metrics.* The Metrics Domain Viewpoint addresses the concern—What types of metrics (i.e., operational, developmental, and programmatic) are required to support the generation of IDSK metric-based tables?

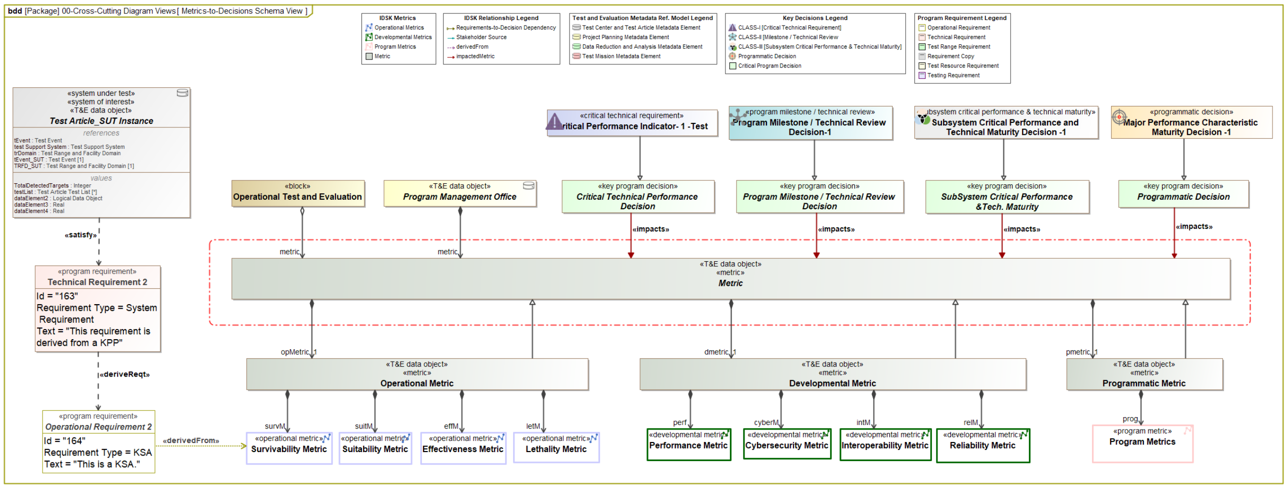


Figure 8. IDSK RA Performance Metrics Domain View

Decisions and Test Article Viewpoints and Views. Figures 9-A and 9-B depict views that portray the IDSK RA from the viewpoints of a decision class and test article with emphasis on the key relationships between these IDSK elements and those that are relevant for the generation of standardized test planning IDSK tables. These Viewpoints and corresponding Views address the concerns—What are the required relationships and structural elements needed to support the generation of the MB-IDSK *test article* and *test resource* standardized views?

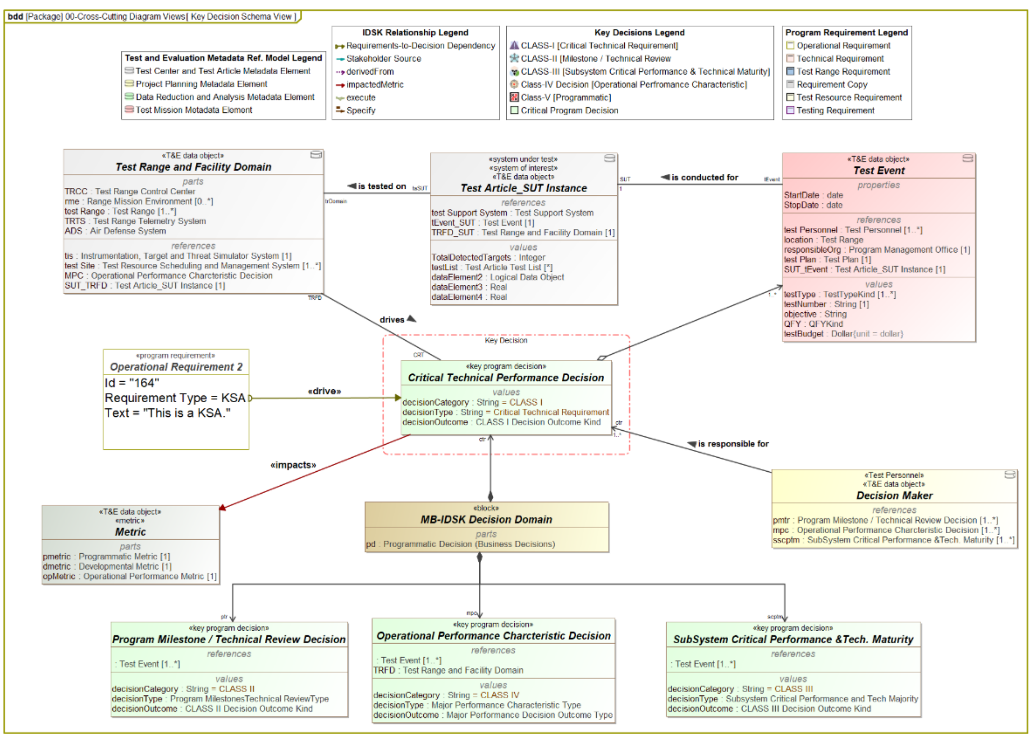


Figure 9-A. IDSK RA Class I Decision-Type View

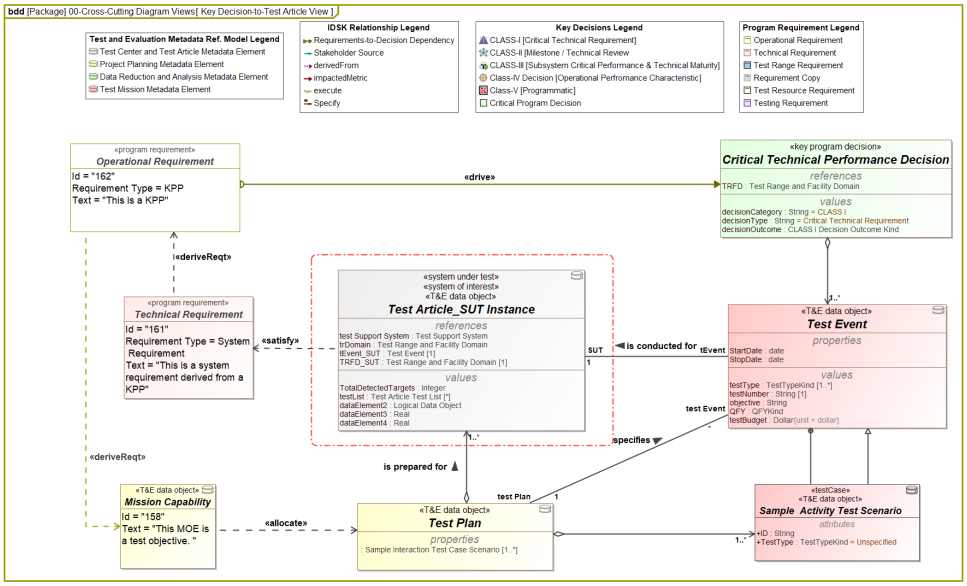


Figure 9-B. IDSK RA Test Article View

Decision-Maker and Test Planning Data Viewpoints and Views. The Decision-Maker and Test Planning viewpoints of the IDSK architecture are created to focus attention on the test planning and decision support needs and concerns of the Decision-Makers (e.g., PO) regarding the T&E of the System-of-Interest (i.e., SOI/SUT). The Decision-Makers within the PO are the primary decision-making authority and are responsible for each decision as illustrated in Figure 5. The PO is comprised of most key Decision-Makers and has ownership of the Test Article and the Decisions that need to be made. Defining the Decision-Maker viewpoint allows views to be created that provide critical insights into the relevant relationships between IDSK elements and how these relationships can be leveraged to support decision-making at each phase of the T&E process. As shown in Figure 10-A, elements specified in to the Decision-Maker Viewpoint include the *PO*, *decisions*, *metrics*, *operational requirements*, *technical requirements*, and *test article* elements respectively.

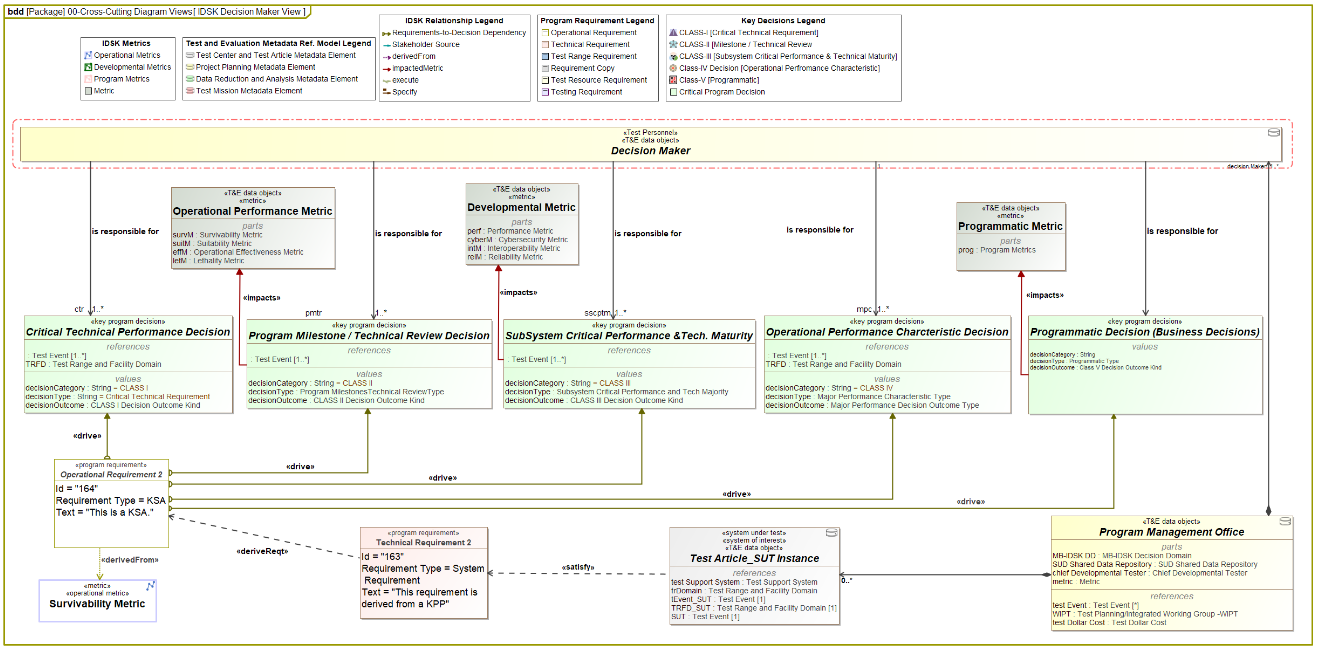


Figure 10-A. IDSK RA Decision-Maker Domain View

Figure 10-B also defines elements that represent data sources relevant to the decision space such as *test range*, *test event*, *test article*, *test personnel*, and elements that capture crosscutting data. Data elements from this view are leveraged in most of the IDSK standardized table views.

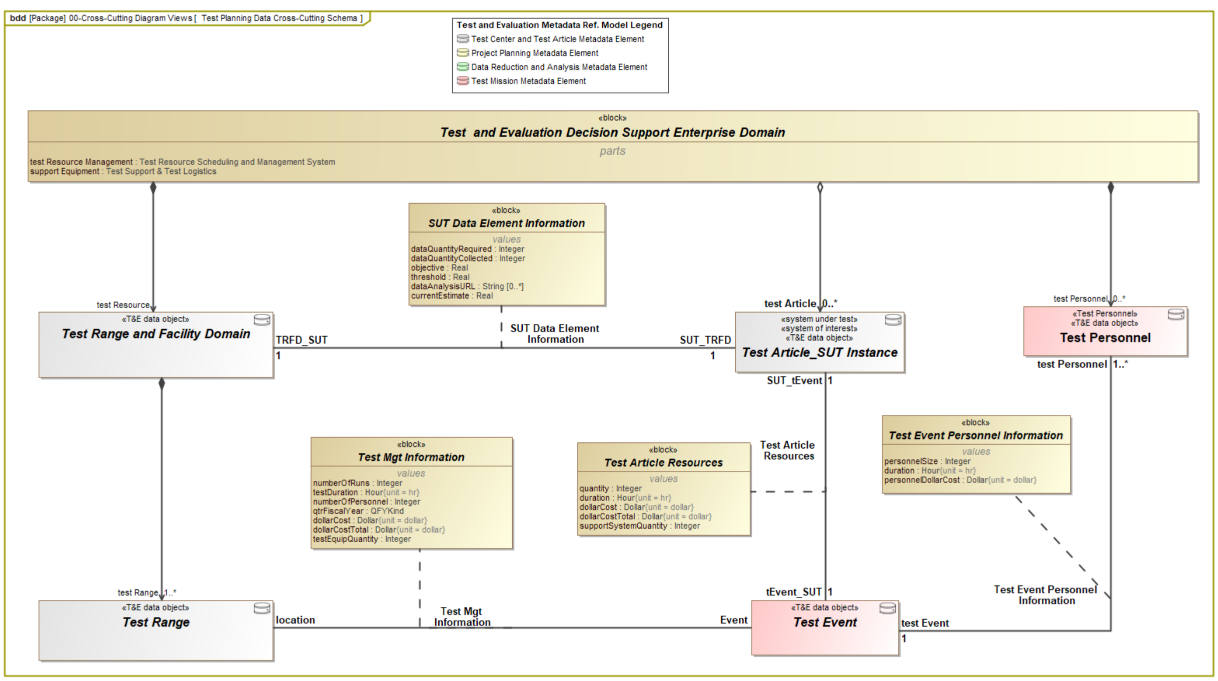


Figure 10-B. IDSK RA Data Sources View

Requirements and Mission Viewpoints and Views. The IDSK RA Requirements view depicted in Figure 11-A portrays various types of Requirements defined as part of the IDSK RA. This architectural view provides insight regarding the IDSK RA's requirements pattern/schema and how each requirement type maps to several architectural elements such as *the test range* *and facility domain*, *test article*, *test case scenario*, *metrics*, and *key program decisions*. As illustrated in Figure 11-A, Technical Requirements are *derivedFrom* Operational Requirements (i.e., KPPs, KSAs) while the Operational Requirements are *derivedFrom* Metrics and *drive* the Key Program Decisions. Specified test range Requirements trace to Operational Requirements and are satisfied by the test range capability required to enable testing of the systems-of-interest. It is important to note that the IDSK RA requirement view shown below does not represent all requirement types needed to support the generation of the IDSK requirements-related tables. The Requirements Viewpoint and View addresses the concern—What are the requirements, relationships, and IDSK elements needed to support the IDSK requirements-related standardized data formats? The IDSK RA Mission view depicted in Figure 11-B defines a few key elements and relationships from a Missions Viewpoint which are required for decision support during acquisition T&E.

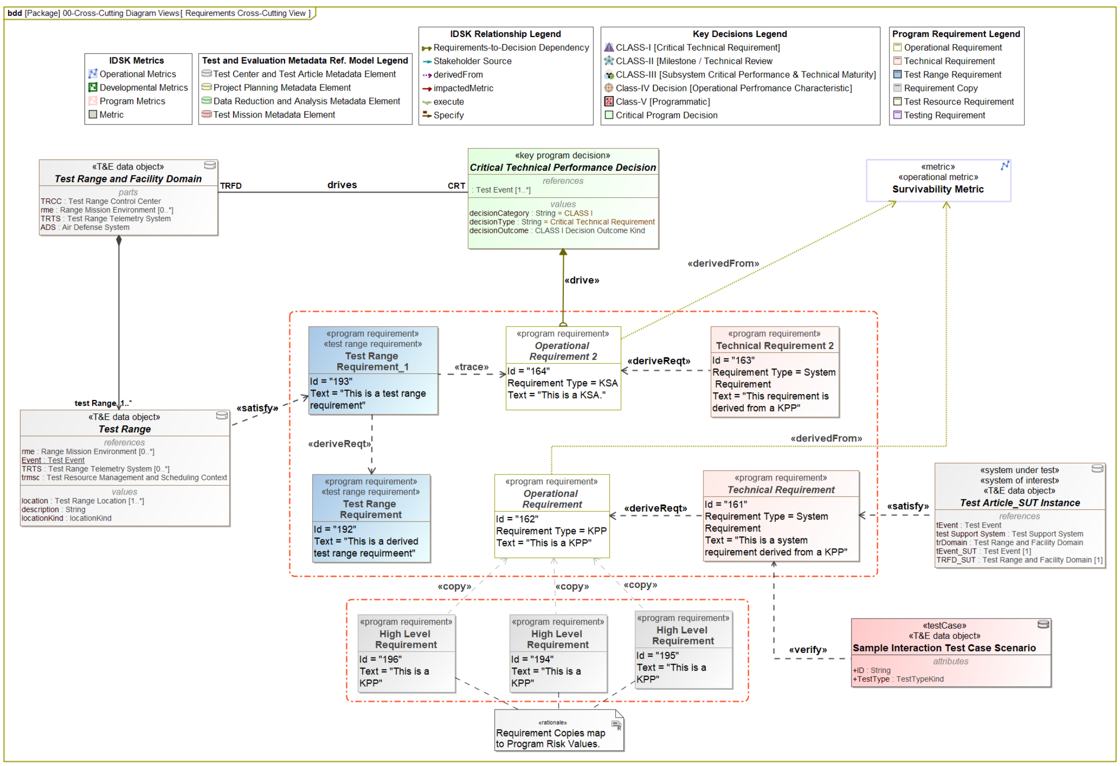


Figure 11-A. IDSK RA Requirements View

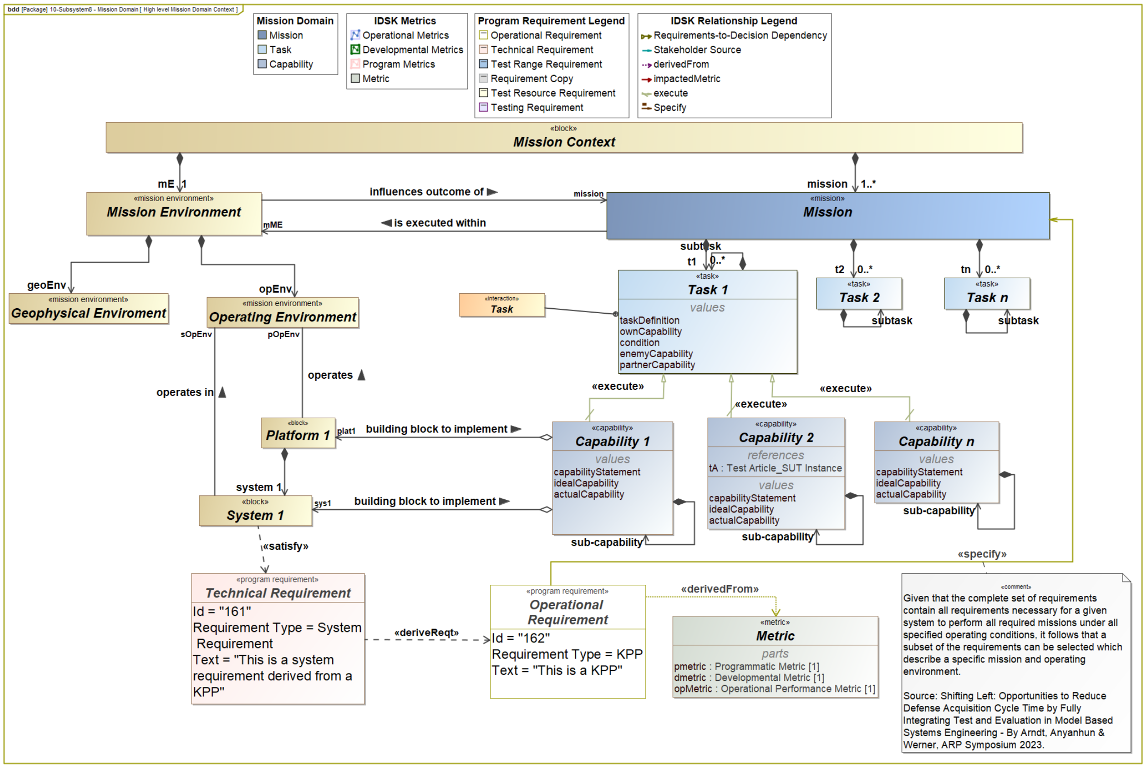


Figure 11-B. IDSK RA Mission View

Program-Specific MB-IDSK Development Process: Instantiating the MB-IDSK RA

A how-to step-by-step architecting process for developing program-specific architectures is captured as part of the MB-IDSK RA model. A high level developmental process view, which outlines the steps a PO utilizing the IDSK RA should take to achieve an MB-IDSK, is portrayed in Figure 12.

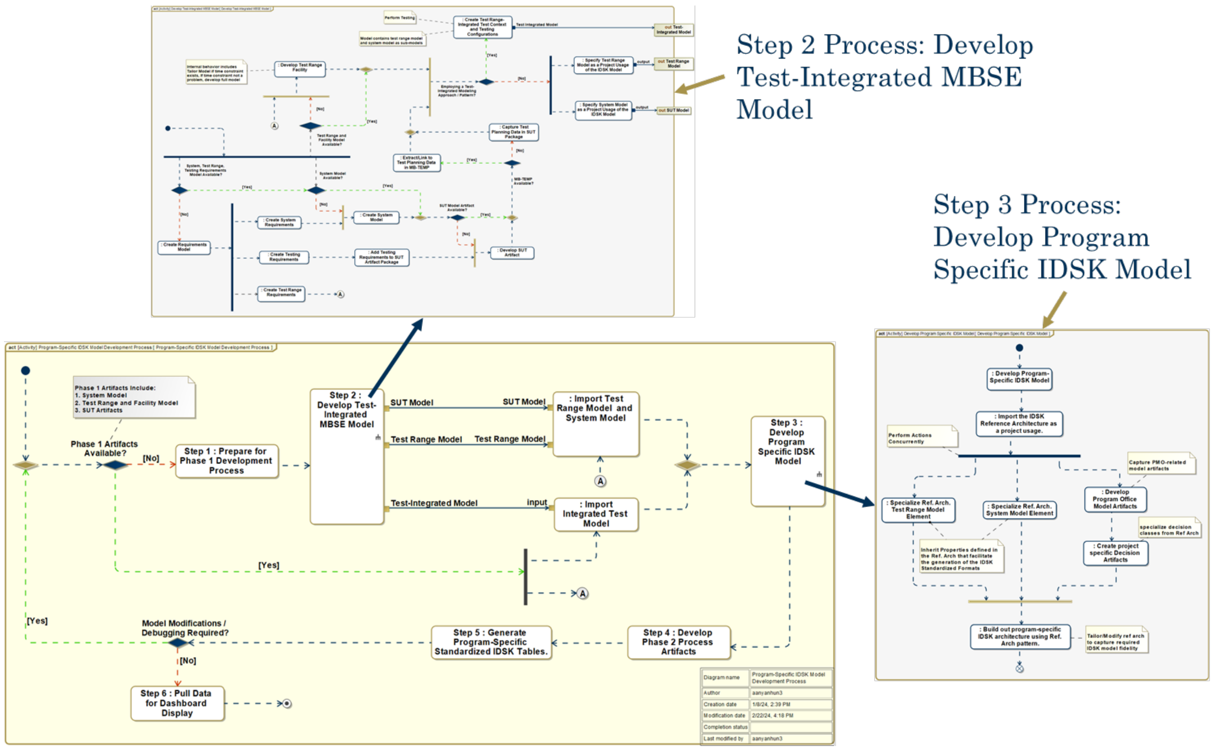


Figure 12. Program-Specific MB-IDSK Development Process

As depicted in Figure 12, the MB-IDSK development process is split into two phases with Phase 1 activities being the development of a program's digital (system) models—system model, requirements model, test model, and so forth. In the case of a program implementing MBSE, most Phase 1 artifacts may already exist, in which case the program IDSK lead need only focus on (1) developing the program-specific IDSK artifacts of Phase 1 and (2) Phase 2 activities, which include generating the standardized IDSK table views.

Figure 13-A describes the model package setup for a program-specific IDSK. As illustrated, the MB-IDSK utilizes data and artifacts from already existing digital (system) models as input for the IDSK. This approach prevents the duplication of data and modeling effort, as well as ensures the integrity and trustworthiness of the data on which decision-makers must depend for making decisions. Although a profile containing stereotypes and customization elements was created as part of the RA to extend the SysML, the use of the inheritance mechanism via the generalization/specialization relationship—shown in Figure 13-B—is the primary means by which concrete implementations realize the properties and relationships already specified in the RA.

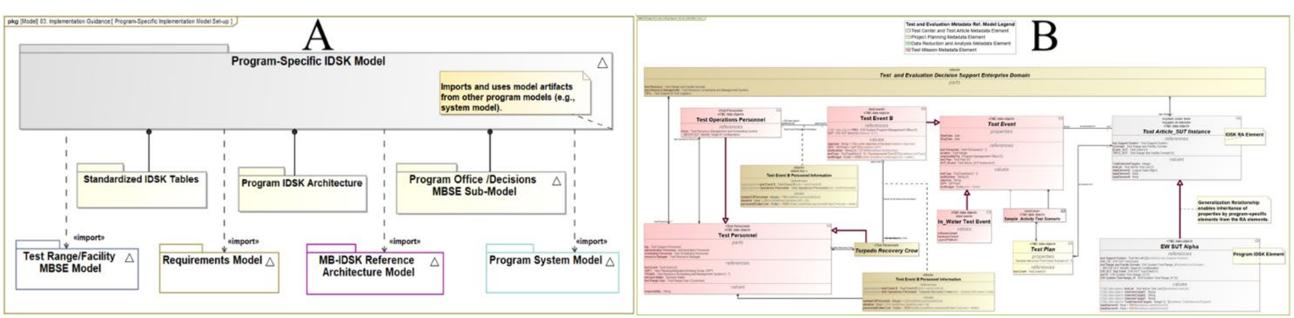


Figure 13-A. Program IDSK Model Setup & Figure 13-B. IDSK RA Instantiation View

Model artifacts developed to assist programs in developing the IDSK include a model library with sets of table templates and query mechanisms as shown in Figure 14-A and 14-B, an IDSK SysML profile, a conceptual and logical data model, a standardized set of tailorable and extendable key decisions (Class I, Critical Technical Requirements; Class II, Program Milestones/Technical Reviews; Class III, Sub-System Critical Performance and Technology Maturity; Class IV, Major Performance Characteristics; and Class V, Programmatic Decisions), and a format for collecting data about these decisions and other enabling resources that together help shorten the architecture development cycle time for program-specific implementations.

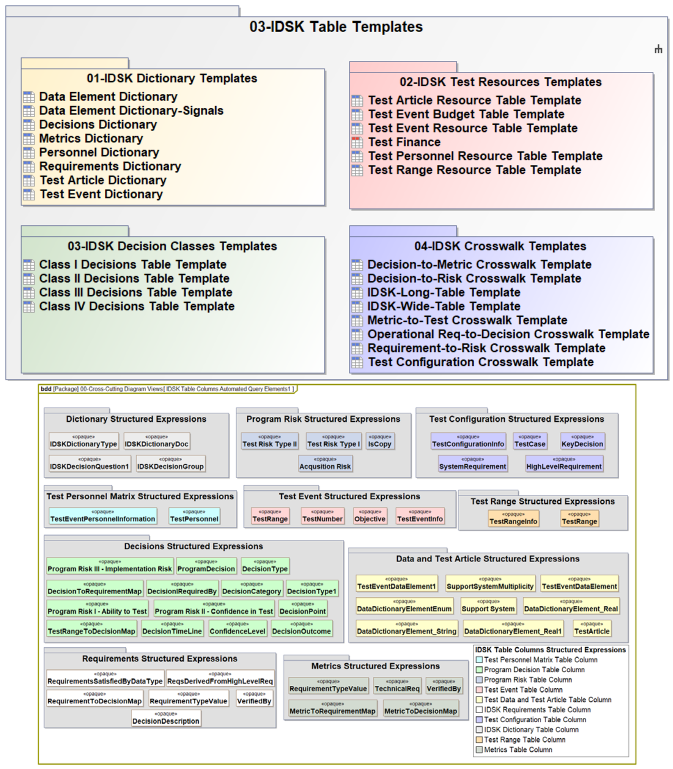


Figure 14-A. IDSK Table Templates & Figure 14-B. IDSK Library of Query Expressions

Exemplar Electronic Warfare System IDSK Architecture and Tables

The architecture view shown in Figure 15 portrays the IDSK decision support domain for an Electronic Warfare (EW) system program developed to support decision-making and test planning for a *Detect Target Id* Test Event. To generate the necessary IDSK standardized tables, the *generalization* relation is used between the more general RA elements and those shown in Figure 15. This modeling approach enables the elements of the EW System T&E Decision Support Domain to inherit and redefine properties and relationships already defined in the IDSK-RA.

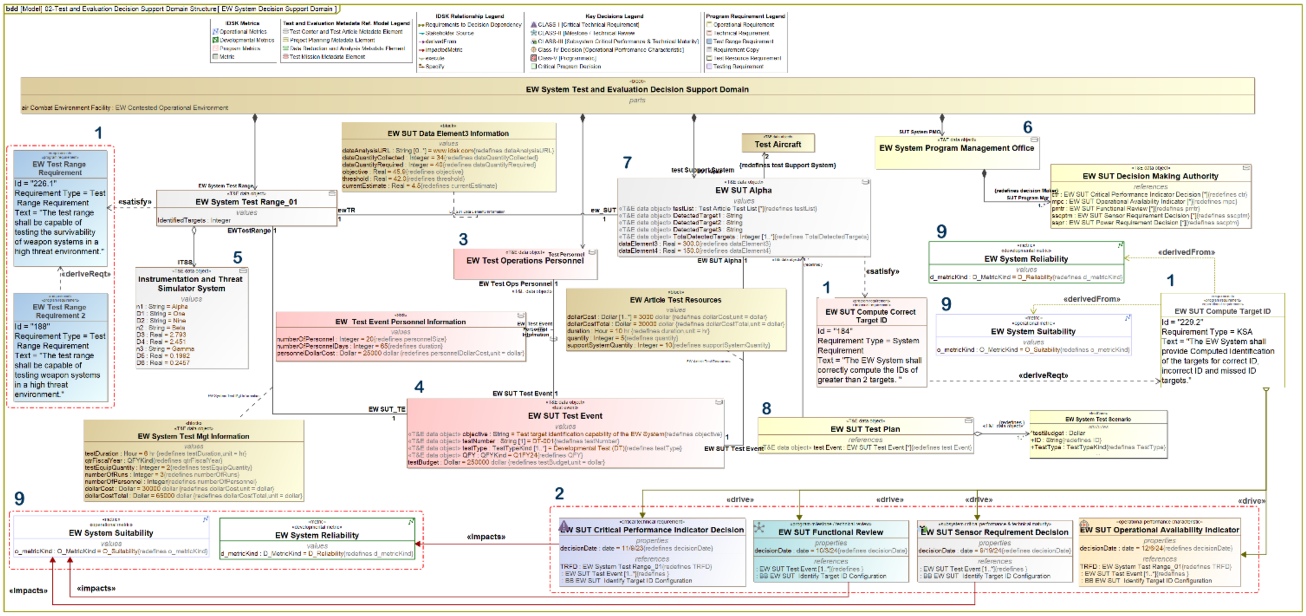


Figure 15. EW System T&E Decision Support Domain View

IDSK architecture elements defined for the EW system exemplar include (1) Requirements, (2) Decisions (Class I–IV), (3) Test Personnel, (4) Test Event, (5) Test Range Instrumentation, (6) Program Office (7), Test Article (EW System), (8) Test Plan, and (9) Metrics. To support the generation of decision support views from a model-based test execution and analysis context, an EW system *Detect Target Id* test context adapted from Arndt et al. (2023) was developed.

The test context was created and used to perform a simple black box test execution of two test case scenarios. Figure 16-A and Figure 16-B describes the testing configuration (1) consisting of the EW SUT, test range, test instrumentation, and test personnel respectively; (2) the SUT behavior modeled using an ACT diagram; and (3) and (4) two test cases executed per test run.

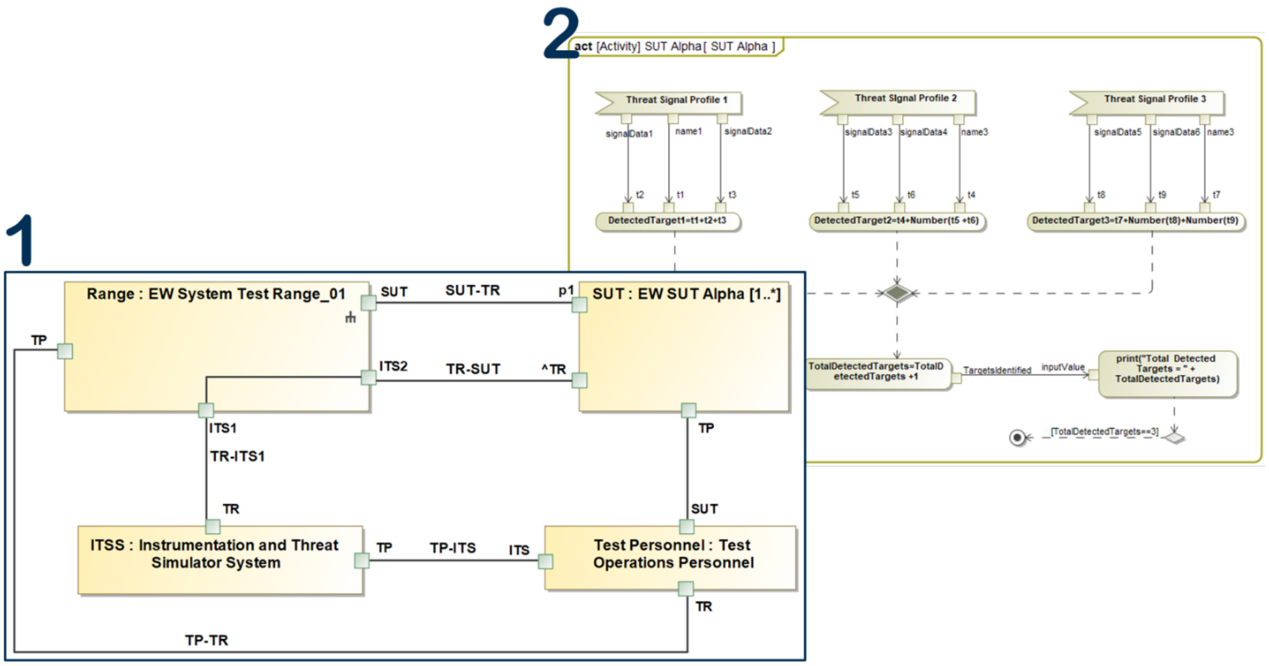


Figure 16-A. EW System Detect Target Id Test Configuration

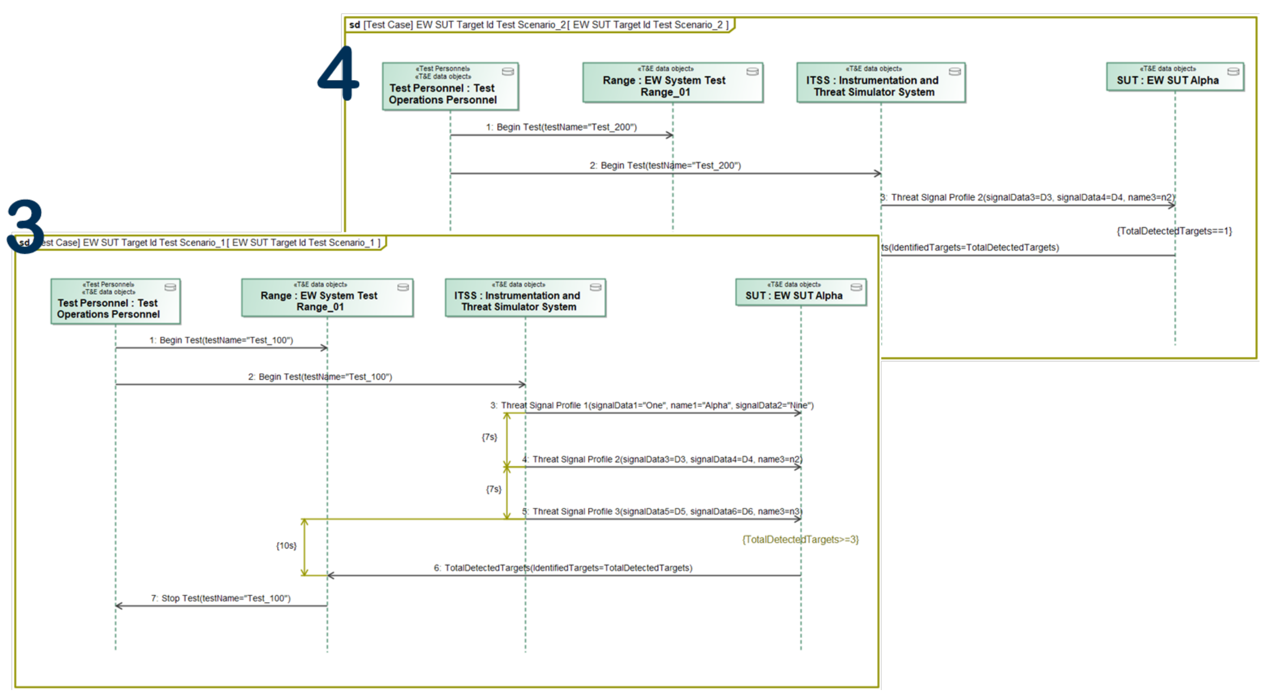


Figure 16-B. EW System Detect Target Id Test Scenarios

EW System *Detect Target Id Test* IDSK Tables. The IDSK data formats portrayed in Table 10–Table 10-15 are generated from the EW System IDSK model using the standardized decision and test planning templates created as part of the IDSK-RA library. IDSK decision views are portrayed in Table 10–12. Table 13 portrays a Test Personnel Resource table, while

Table 14 and Table 15 illustrate an exemplar IDSK wide table and a Test Configuration Crosswalk table respectively.

Table 10. Detect Target Id Class I Decision Crosswalk Table (Notional)

#	Name	Program Decision	Decision Type	Decision Category	Decision Outcome	Decision Date	Lifecycle Point	Confidence Level Required for Decision	Operational Requirement	Technical Requirement	Data Required	Data Source
1	EW SUT Critical Performance Indicator Decision	Decision question goes here	Critical Technical Requirement	CLASS I	Inconclusive	11/9/23	Milestone B	Unspecified	229.2 EW SUT Compute Target ID	184 EW SUT Compute Correct Target ID	Total Detected Targets	Test

Table 11. Detect Target Id Class II Decision Crosswalk Table (Notional)

#	Name	Program Decision	Decision Category	Decision Type	Confidence Level Required for Decision	Lifecycle Point	Decision Date	Decision Outcome	Operational Requirement	Technical Requirement	Data Source	Data Required
1	EW SUT Functional Review	DOES THE DESIGN MEET THE FUNCTIONAL REQUIREMENT NEEDS?	CLASS II	Milestone A - Technology Development and Risk Reduction	High Confidence (> 90%)	Critical Design Review	10/3/24	Pass Review	229.2 EW SUT Compute Target ID	184 EW SUT Compute Correct Target ID	Analysis	Data required goes here

Table 12. Detect Target Id Class IV Decision Crosswalk Table (Notional)

#	Name	Program Decision (Class IV)	Decision Category	Decision Type	Confidence Level Required for Decision	Lifecycle Point	Decision Date	Decision Outcome	Operational Requirement	Technical Requirement	Applicable Test Range	Test Data Required	Data Source
1	EW SUT Operational Availability Indicator	WILL THE SYSTEM MEET OPERATIONAL AVAILABILITY NEEDS?	CLASS IV	Operational Availability	High Confidence (> 90%)	Critical Design Review	12/6/24	Unspecified	229.2 EW SUT Compute Target ID	184 EW SUT Compute Correct Target ID	EW System Test Range_01	Data Required type goes here	Test

Table 13. Detect Target Id Test Personnel Resource Table (Notional)

#	Name	Test Personnel Type	Number of Test Personnel	Duration (Hours)	Personnel Dollar Cost
1	EW SUT Test Event	EW Test Operations Personnel	20	65	25000

Table 14. Detect Target Id IDSK Wide Table (Notional)

#	Name	Metrics	Operational Requirement	Data Element 3	Data Element 4
1	EW SUT Sensor Requirement Decision	EW System Suitability	229.2 EW SUT Compute Target ID	300.0	150.0

Table 15. Detect Target Id Test Configuration Crosswalk (Notional)

#	Name	TestResult1 : VerdictKind	TestResult2 : VerdictKind	SUT : EW SUT Alpha	SUT.TotalDetectedTargets : Integer	Test Personnel : Test Operations Personnel	Test Event_1.testType : TestTypeKind	Test Event_1. testNumber : String	D1 : EW SUT Critical Performance Indicator Decision	D3 : EW SUT Sensor Requirement Decision	Range : EW System Test Range_01
1	EW System Test	pass	fail	sut[1] : EW SUT Alpha sut[2] : EW SUT Alpha	3 1	: Test Operations Personnel	Developmental Test (DT)	1001	: EW SUT Critical Performance Indicator Decision	: EW SUT Sensor Requirement Decision	: EW System Test Range_01

Conclusion

The pivot to a digital engineering approach for IDSK development through the use of MBSE accelerates the delivery of data needed to inform acquisition and T&E decision-making. The MB-IDSK RA approach presented in this paper provides decision support in the form of standardized decisions and test planning data formats, to adequately equip decision-makers with data needed to inform critical decisions. As a decision support tool, the MB-IDSK RA pulls/aggregates data from other digital (system) models within a program's digital engineering ecosystem to equip decision-makers with timely pertinent data from trusted data sources required to make the best decisions. Primarily, the MB-IDSK RA is a lightweight RA created to foster flexibility and evolvability as its key quality attributes to ensure it is easily realizable, adaptable, and can guarantee its usefulness and practicality to program offices and the T&E enterprise. Moreover, the RA enables the development of tailorable program-specific architectures from which IDSK table views can be realized. Specifically, the table formats generated using the IDSK RA include tables that may be classified as either an IDSK Dictionary, IDSK Resource, IDSK Crosswalk, or IDSK Decisions table respectively. Additionally, a standardized set of program decisions and a format for collecting data about these decisions are developed as part of the IDSK RA. Consequently, the approach to decision support and test planning demonstrated in this work is a critical missing link in the race to deliver advanced systems to warfighters at the speed of need. Most importantly, it facilitates accelerated delivery of T&E data to decision-makers to inform decision-making.

Future Work

The adoption of MBSE by a wide range of DoD programs has led to a number of significant improvements in the acquisition development lifecycle. The development of the MB-IDSK RA is a great example of these improvements. Notwithstanding, although the IDSK RA allows for the specification of complex relationships between decisions, data, testing, and a number of different program elements—for decision-makers—the complexity of an MB-IDSK could be a problem. An additional challenge within the existing DoD workforce is the apparent lack of MBSE modelers with the requisite skillset and expertise required to create, populate, and maintain an MB-IDSK. To make the different aspects of the complex multidimensional relationships easier for decision-makers to understand, additional work needs to be done in the development of visualization tools. Furthermore, organizations like program offices tasked with the responsibility to develop the MB-IDSK would benefit from simple data entry utilities that would enable programs and T&E personnel with little understanding of SysML models to simply populate the different parts of the MB-IDSK.

References

Army Aviation and Missile Command Fort Eustis, VA. (2022). *Comprehensive architecture strategy (CAS)*. https://apps.dtic.mil/sti/pdfs/AD1185001.pdf

Arndt, C., Anyanhun, A., & Werner, J. S. (2023). *Shifting left: Opportunities to reduce defense acquisition cycle time by fully integrating test and evaluation in model based systems engineering*. Acquisition Research Program.

Beers, S. M. (2022, July). *Integrated decision support key (IDSK) for capability evaluation-informed MDO decision-making* [ITEA MDO Workshop].

Beers, S. M., Hutchison, S., & Mosser-Kerner, D. (2013). Developmental test evaluation framework: Focus on evaluation and analysis for acquisition success. *Phalanx*, *46*(3), 36–39.

Collins, C. & Beers, S. M. (2021, December). *Mission engineering, capability evaluation & digital engineering informing DoD technology, prototyping and acquisition decision* [NDIA S&ME Conference].

Executive Services Directorate. (2020). *Test and evaluation* (DoD Instruction 5000.89). Department of Defense. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500089p.PDF

Guertin, N. (2022). *DOT&E strategy update 2022*. https://www.dote.osd.mil/Portals/97/pub/reports/FINAL%20DOTE%202022%20Strategy%20Update%2020220613.pdf?ver=KFakGPPKqYiEBEq3UqY9lA%3d%3d

ISO/IEC/IEEE. (2022). *Systems and software engineering—Architecture description* (SO/IEC/IEEE 42010:2022[en]).

Muller, G. & Hole, E. (2007, March). *Reference architectures; Why, what and how* [White paper]. Architecture Forum Meeting, Embedded Systems Institute and Stevens Institute of Technology, Hoboken, NJ, United States.

Werner, J. S., & Arndt, C. (2023). *Development of digital engineering artifacts in support of MBSE-based test planning, execution, and acquisition decision making*. Acquisition Research Program.

Accelerating Implementation of Critical Joint Warfighting Concepts and Capabilities

Dr. Laura Freeman—is the Deputy Director of the Virginia Tech National Security Institute and is the Assistant Dean of Research for the College of Science. Freeman previously served as the Assistant Director of the Operational Evaluation Division at the Institute for Defense Analyses. She served as the Acting Senior Technical Advisor to the Director, Operational Test and Evaluation in 2018. She has a B.S. in Aerospace Engineering, M.S. in Statistics, and Ph.D. in Statistics, all from Virginia Tech.

Dr. Sandra Hobson—is a Senior Executive in the Department of Defense responsible for the development of policy, tools, methods, and capabilities needed for the test and evaluation of operational performance of weapon systems and emerging technologies. She holds a B.S. in Aerospace Engineering from the United States Naval Academy and a Ph.D. from the University of Maryland. Hobson has been honored with a UNESCO National Fellowship for Exceptional Researcher and two Secretary of Defense Medals for Meritorious Civilian Service.

Dr. James Moreland Jr.—retired as a Senior Executive from the Department of Defense in 2020 and is now Vice President of Mission Integration at Raytheon Technologies. Prior to that, he served as Executive Director of Mission Engineering and Integration and Director of Naval Warfare in the Office of the Secretary of Defense and as Chief Engineer of the Naval Surface Warfare Center, Dahlgren Division. He received a B.S. in Mechanical Engineering from the University of Maryland, an M.S. in Systems Engineering from Virginia Tech, an M.S. in National Resource Strategy from the Industrial College of the Armed Forces, and a Ph.D. in Systems Engineering from The George Washington University.

Mr. Scott Lucero—retired from the Office of the Secretary of Defense in 2020 and is now a Senior Research Associate at the Virginia Tech National Security Institute. He served in a variety of positions reinvigorating DoD's practice of systems engineering in the Office of the Secretary of Defense before retiring as the Director of Systems Engineering Transformation. He also served 8 years as lead for software test and evaluation in the Army's Test and Evaluation Command. He is a Ph.D. candidate in Industrial and Systems Engineering at Virginia Tech and received B.S. and M.S. degrees in Computer Science from Old Dominion University and The George Washington University.

Dr. Jeremy Werner—is currently the Science Advisor in the Office of the Director, Operational Test and Evaluation. Prior to that, he served at Johns Hopkins University Applied Physics Laboratory, where he founded a data science–oriented military operations research team that transformed the analytics of an ongoing military mission. He also served at the Institute for Defense Analyses, where he supported the rigorous assessment of a variety of systems/platforms. He holds a Ph.D. in Physics from Princeton University as well as a B.S. in physics from University of California, Los Angeles.

Dr. Raymond O'Toole—has more than 30 years of experience as a naval officer (active and reserve), retiring at the rank of captain. He is currently the Principal Deputy Director, Operational Test and Evaluation serving as the principal staff assistant to the Director for all functional areas assigned to the office. O'Toole holds a B.E. in Marine Engineering. He also holds an M.E. in Systems Engineering from Virginia Polytechnic Institute and State University, an M.S. in National Resource Strategy from the Industrial College of the Armed Forces, and a Ph.D. in Engineering Management from The George Washington University, where he is now a Professional Lecturer of Engineering Management and Systems Engineering. He is a recipient of two Secretary of Defense Meritorious Civilian Service Awards and the Department of the Navy Meritorious and Superior Civilian awards.

Abstract

Mission engineering is a relatively new discipline born out of the need to support mission planners and strike authorities with emerging technologies and innovative solutions to achieve mission success in complex, multi-domain operating environments. Mission engineering combines the future operating environment and the strategic intent outlined in the National Defense Strategy with the rigor of system engineering, software engineering, digital engineering, and related disciplines, to identify critical operational gaps and architect the system of emerging materiel and

non-materiel solutions required to reach the desired strategic or tactical objectives. The ultimate objective is to optimize mission accomplishment and outcomes by advancing the existing operation plans, kill-webs, mission threads, and vignettes with innovative technologies and capabilities to deter or defeat any adversary in the most complex engagements. This article discusses the challenges of mission engineering and proposes the integration of operational and live fire test and evaluation within this process to mitigate some of those challenges.

Introduction

> Organizing to secure our advantage is not just a strategic goal; it is an imperative that assures our nation's future defense. No warfighting domain remains uncontested. The complexity of warfighting is growing with technology, so no single Service capability can win alone without truly realized joint force capabilities.

—Admiral Christopher W. Grady, "Sharpening Our Competitive Edge: Honing Our Warfighting Capabilities Through the Joint Warfighting Concept"

Mission engineering that is grounded in test data and accredited modeling and simulation (M&S) results is one of many tools that identify gaps and look for creative, out-of-the-box solutions to respond to the persistent adversary and dominate in such contested environments. Many other never-before-seen technologies and tactics, techniques, and procedures (TTPs) may be delivered because of rigorous mission engineering efforts that could otherwise not have been identified if the problem was viewed purely at a system level, or even at an individual Department of Defense (DoD) component level.

The current practice of mission engineering relies on M&S—high-level campaign analyses—with the capability to run what-if scenarios to identify gaps in achieving mission success for different vignettes. Mission engineering analyses could guide the department's decisions regarding distributed maritime operations and dynamic expeditionary operations in a high-end-fight, employing thousands of attritable, fully autonomous systems to overwhelm the adversary in mass and achieve the desired lethal effect. Mission engineering would consider various alternatives in the context of the ability to establish electromagnetic spectrum superiority and the possibility that the friendly communications' emissions present a significant susceptibility to being targeted by the enemy. It is necessary to account for these synergistic and emergent effects and the array of possibilities, shown in Figure 1, to realistically represent and exploit the highly complex battle space and achieve an enduring advantage. Mission engineering involves forecasting the performance of future capabilities to inform future requirements and acquisition priorities that will in turn drive science and technology investments. Because of its heavy reliance on M&S and its focus on the operational performance and mission success, it is important to integrate mission engineering efforts with the operational and live fire test and evaluation activities. This integration involves not just the operational and live fire testing of individual systems but also such testing of joint warfighting concepts, kill-webs, vignettes, mission threads, and other system-of-systems scenarios. The integration of these two disciplines can enhance the realism of the mission engineering architecture by identifying and supplying relevant operational data critical to verifying and validating the mission engineering outputs. This article illuminates and explores the synergistic benefits of this integration.

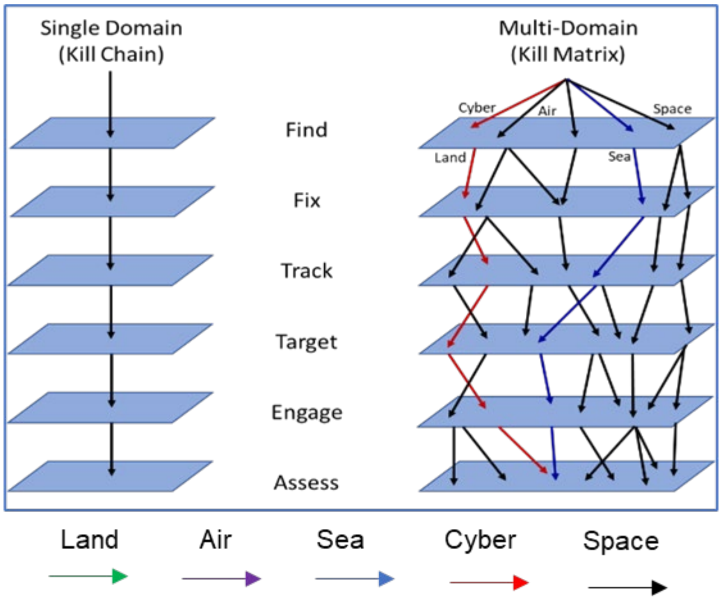


Figure 1. Kill-Web Combinatorics

Mission Engineering

Mission engineering is an interdisciplinary process encompassing the entire technical effort to analyze, design, and integrate current and emerging operational needs and capabilities to achieve desired mission outcomes (Office of the Under Secretary of Defense for Research and Engineering [OUSD(R&E)], 2023). Mission engineering decomposes missions into their constituent parts to

- identify gaps, challenges, and opportunities
- inform decisions regarding requirements, architectures, and technologies needed to achieve the combatant commanders' strategic and tactical mission objectives

The Under Secretary of Defense for Research and Engineering developed a five-part process for mission engineering, shown in Figure 2 (OUSD[R&E], 2023):

1. Frame the mission problem or opportunity.
2. Characterize the mission, including specific scenarios, vignettes, and measures.
3. Model the mission architectures.
4. Perform analysis and evaluate trade-offs.
5. Document results and recommendations.

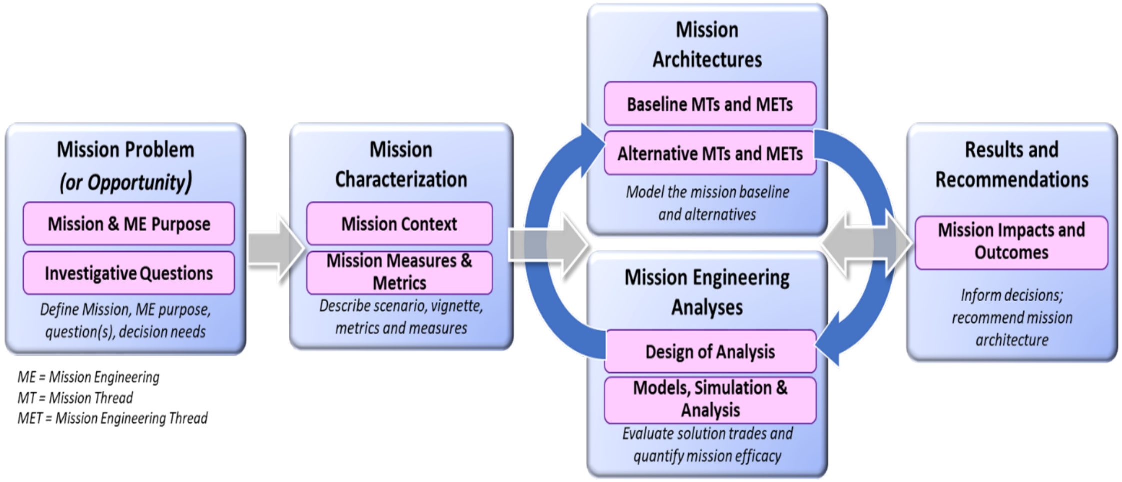


Figure 2. Elements of the Mission Engineering Process

Operational Test and Evaluation in the Context of Mission Engineering

Operational and live fire testing determine the operational effectiveness, suitability, survivability, and lethality of the systems and services that the DoD acquires through the Defense Acquisition System. By Section 139, 4171 and 4172 of Title 10, United States Code, and, more elementarily, by the scientific method that the law is written to reinforce, such determinations cannot be based solely on M&S. They require live data collected in operationally representative conditions using trained operators, maintainers, and defenders. While mission engineering is not intended to determine the operational performance of the to-be fielded systems, it does inform other equally important decisions in support of the warfighter. As the DoD moves into more complex warfighting domains, mission engineering architectures would benefit significantly from being shrewdly rooted in M&S that has undergone rigorous verification and validation using operational and live fire test and evaluation data. This is discussed next in more detail in the context of the five-part process for mission engineering.

Mission Problem

One application of mission engineering is determining the optimal mix of forces to achieve the desired mission effects while expending the least number of resources (Brown et. al., 2023). While operational testing does not typically address this topic, the knowledge gleaned from operational testing may have resource impacts. For example, operational testing may identify operational effectiveness limitations showing that more weapons are required to achieve the desired strategic or tactical effects on the intended targets than originally estimated in mission engineering. Similarly, operational testing may identify operational suitability limitations showing that more systems are required to achieve the same mission effects due to reliability, availability, or maintainability shortfalls than originally estimated using mission engineering. Lastly, live fire testing might identify survivability limitations that again show additional resources are required to account for potential kill removals than originally estimated in mission engineering. One of the chief differences between mission engineering and operational and live fire test and evaluation is that the former focuses on the optimal means to accomplish a mission, while, at least historically, the latter focuses on the mission performance of individual systems. This difference points to an opportunity to use mission engineering to design operational and live fire tests to evaluate future joint warfighting concepts, kill-webs, vignettes, mission threads, and other system-of-systems scenarios—termed in this article **Joint Test**

Concept—to support the collection of definitive data sources underpinning the credibility of mission engineering outcomes.

Mission Characterization

The mission characterization describes the set of variables that provide the context for, among others, the mission objectives, environment, friendly and enemy forces, timeframe, assumptions, constraints, and TTPs. Mission scenarios or vignettes may be derived from joint warfighting concepts, operational plans, concepts of operations, and other mission plans. While mission engineering characterizes the scenarios and mission threads—including the order of battle, threats, and rules of engagement—to identify gaps and solutions and optimize mission outcomes, operational testing characterizes the ability of the system to either execute or contribute to that mission. Mission engineering digital environments enable the evaluation of a broader set of mission contexts, which could inform plans for operational or live fire test and evaluation—especially if future opportunities include the evaluation of the operational performance of vignettes, kill-webs, mission threads, and the like.

Mission Architecture

Mission architecture models the concepts, approaches, and full system of systems to examine the entire mission's process and data flow, interactions and timing, and capabilities and performance required to meet the mission objective. Mission threads are the elements of this architecture that describe the various assets and end-to-end tasks needed to accomplish a specified mission. A mission engineering thread assigns systems, organizations, or assets to perform a task as shown in Figure 3. These mission threads are available from the Joint Staff's Universal Joint Task Lists. Based on the mission thread analysis, operational or live fire test plans could identify the requisite live data and accredited M&S results needed to validate the mission engineering outcomes. The integration of these two disciplines could help ensure that the results of each will provide deeper insight into different aspects of the identified challenges. For example, sharing a common representation of the threat and TTPs could ensure alignment in several areas, including the format and fidelity of threat surrogate digital artifacts.

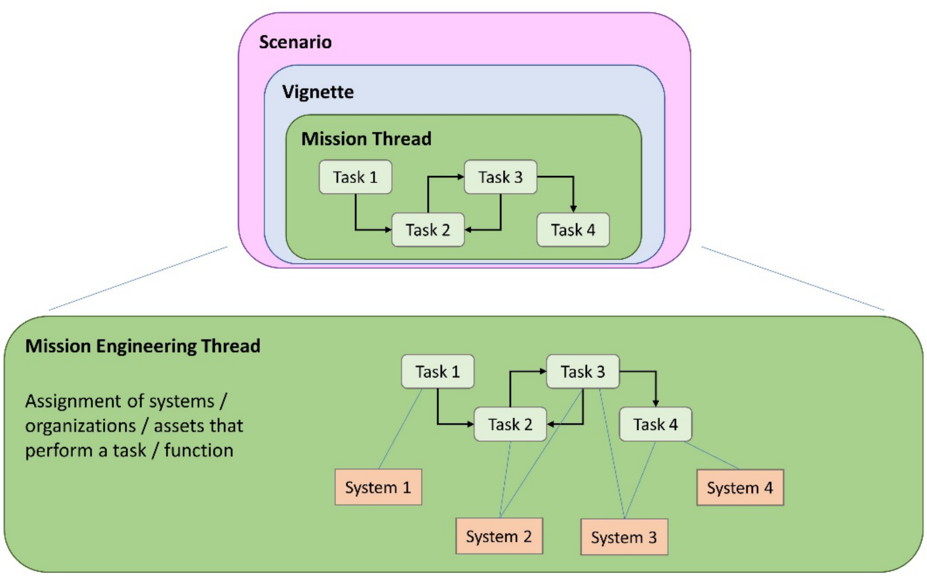


Figure 3. Mission Threads and Systems (Mission Engineering Guide, 2020)

Mission Analysis

Mission engineering and operational test evaluate mission success in different ways. Mission engineering forecasts mission scenario outcomes using M&S and sensitivity analyses to understand how uncertainty propagates across the M&S, while operational testing measures an individual system's contribution to a mission scenario using a combination of live data and accredited M&S results. For example, operational testing measures operational availability and logistics delays, which can play an important role in determining mission outcomes, especially for sustained operations. Both mission engineering and operational testing frequently use force exchange ratios in their evaluation of overall mission effectiveness. The commonality in evaluation areas offers an opportunity for mission engineering to integrate and inform operational system performance measurements collected in operational and live fire test and evaluation. Conversely, mission engineering can extend operational and live fire system performance measurements into a wider range of mission contexts. Mission engineering also often forecasts acquisition and sustainment costs, return on investment, or other cost–benefit–related quantities, whereas operational testing typically does not, although the operational performance demonstrated in test can inform those analyses.

Results and Recommendations

In addition to documenting the analysis results and recommendations, mission engineering also calls for development of a preferred mission architecture and curation of the data, models, and architectures used to produce the results. Operational and live fire test and evaluation reports document the adequacy of the testing that was planned and executed to determine operational performance. They also report on the operational effectiveness, suitability, survivability, and lethality (as applicable) of the system in operationally representative conditions. Lastly, they report on recommendations to address any deficiencies in observed system performance or test limitations that precluded the evaluation of some aspects of the system performance. Both disciplines could leverage each other's reports and data, but this would warrant the development of coordinated data strategies outlining data curation, analysis, and storage needs. These data curation activities are essential to developing a common understanding of mission and system performance and developing realistic assessments of mission success, fully realizing the benefits of each activity.

Opportunities for Leveraging Operational and Live Fire Test Data to Enhance Mission Engineering Outputs

Curiously, the *Mission Engineering Guide* mentions the term *data* more than 60 times but *system testing* just three. The guide does, however, state, "For the purposes of mission engineering, the term "data" means information related to the scenario or vignette, OOB [order of battle], force structure, system parameters or performance, threat, models, and analytical results" (OUSD[R&E], 2023).

This statement implicitly suggests that operational and live fire test data of DoD *systems*—but also of mission *scenarios and vignettes*—should be the basis upon which mission engineering becomes data-driven and more realistic. Focusing operational and live fire testing on only one system may not capture all intricacies of the real-world mission scenarios involving the use of multiple systems of varying complexities and pedigrees working together to achieve the desired lethal effect. The emergence of highly network-centric concepts, greater dependency on connectivity, and the use of large amounts of data from a wide array of shooters and sensors across multiple domains, at machine speeds, warrants commensurate operational and live fire test and evaluation. Evaluating warfighting capability is further challenged by asynchronous updates and continuous evolution of the various components that comprise these system-of-systems operations. This complexity demonstrates an inherent need to continually characterize the interoperability of such systems and their effectiveness as employed by the combatant commands. With the emergence of mission engineering, joint all domain command and control solutions, and the concept of kill-webs, it is important to have operational and live fire test and evaluation also effectively measure the success rates of joint warfighting concepts, kill-webs, mission threads, and other system-of-systems solutions. Ongoing DoD efforts are investigating the feasibility of these activities under the Joint Test Concept initiative.

Joint Test Concept

Initial studies have validated the need to revamp traditional operational and live fire test and evaluation to focus on the operational and mission context in which the system under test is expected to perform throughout the system life cycle. The resultant Joint Test Concept initiative further investigates how operational and live fire test and evaluation could be transformed to either leverage existing exercises and experiments or establish a complementary process by which the department can evaluate the *lethality, suitability, resilience, survivability, agility, and responsiveness* of the joint force. The Joint Test Concept considers an end-to-end capability life cycle approach—anchored in mission engineering and a digital environment—calling for a more holistic yet dynamic and flexible approach to assess system performance across three overlapping layers:

- *System Performance Layer*, where the system is evaluated in isolation
- *Capability Immersion Layer*, where the system and mission threads are evaluated in pre-defined systems of systems
- *Joint Capability Demonstration Layer*, where the system is evaluated in a joint multi- or all-domain environment

Joint Test Concept: Leading Tools and Practices

There are key leading practices that may enhance the implementation of the joint test concepts across the three identified layers: (1) M&S, including those used by mission engineering; (2) test infrastructure and networking; and (3) data and artificial intelligence (AI).

Modeling and Simulation

Organizations across the DoD are developing policies and strategies to move forward with implementing digital engineering. A leading practice used with digital engineering is modular open system architecture (MOSA). MOSA approaches encourage interoperability and more rapid integration of capabilities throughout the system life cycle by using open system standards and architecture modularity. Systems designed with a MOSA approach, avoiding vendor and solution-specific interfaces, are designed to more easily integrate and test for joint missions. As emergent threats and new missions illuminate additional joint use cases that may not be part of original system designs, a MOSA backbone will make joint integration more feasible to implement and test, supporting enhanced integration of joint operational and live fire testing—at the mission scenario level—with mission engineering.

Another leading practice within digital engineering is model-based system engineering (MBSE). MBSE models that define system interfaces and functionality can be a critical asset to help the joint test concepts capture and understand how various systems should integrate and function together as a whole. System-of-systems models of the enterprise architecture can be used to better define the combined joint mission, identify joint test cases, and illuminate joint test gaps. A continued focus on MOSA implementation will enable programs to better respond to evolving threats by being able to swap out and upgrade components more easily across the system life cycle. As more programs throughout the DoD successfully implement MBSE and MOSA approaches, the joint test concepts can leverage these designs to facilitate more effective evaluation of joint interoperability and mission scenario success.

Digital twins offer the capability to model and simulate a system's physical, digital, and functional characteristics in a digital format, enabling testing to shift left—all the way to mission engineering—in the product life cycle. Digital twin technology is a key enabler in M&S as the DoD moves toward a Live Virtual Constructive (LVC) testing approach that blends traditional live simulations with virtual and constructive environments. This approach enables more realistic, effective, and affordable joint testing and training environments that are difficult and prohibitively expensive to test in a purely live test format. Due to the complexity of the joint mission, LVC environments are seen as an essential piece of joint testing concept with the potential to enable large-scale joint testing events that are integrated across the DoD components and distributed across multiple geographic locations. Joint LVC environments, such as the Joint Simulation Environment, are enabled by a growing digital engineering backbone across DoD programs and offer significant opportunities to improve joint training, testing, and mission engineering of the future. Improvements to network connectivity and integration are a key enabler to facilitate LVC capabilities that are integrated across multiple DoD ranges offering the opportunity to dramatically reshape how joint testing can be executed in the future.

Test Infrastructure and Networking

As the threat landscape and joint missions continue to evolve, test infrastructure must also continue to evolve to meet the specific needs of newer technologies. Emerging mission sets—including hypersonics, space operations, autonomous systems, and electromagnetic spectrum operations, to name a few—all require new infrastructure to fully support testing at a joint level and scale. Operational test and training infrastructure could construct realistic training and testing environments that are integrated to provide warfighter training across distributed sites, providing the environment to collect mission-level operational data (Marler at al., 2022).

For example, the Joint Integrated Test and Training Center (JITTC) is intended to be the first facility to allow Air Force, Navy, and international pilots to fly integrated live and simulation missions together. The facility plans to link live aircraft tracking data over the Joint Pacific–Alaska Range Complex with simulators inside the JITTC. The JITTC is planned to be "first center capable of joint and multinational force training," providing the capability to "blend synthetic and live-fly training while focusing on training events specific to employment of tactical joint assets" (Air Force, 2023). This could serve as an excellent source of operational test data in support of mission engineering.

The Space Force is also investing heavily in its National Space Test and Training Complex (NSTTC) to build a virtual testing and training environment for space missions that are impossible to physically test on the ground (Albon, 2022). The NSTTC aims to build digital environments to represent satellites' behavior under different operational conditions across a variety of space missions. The NSTTC also plans to include ground and space-based instrumentation, command and control support, and a dedicated cyber test range. In its NSTTC vision document, Space Force (2022) identified joint applicability as one of four focus areas, highlighting the need to support development of joint multi-domain operating concepts and integrate joint mission partners. All test and training complexes could support the collection of operational test data in support of mission engineering objectives.

Investments in computing infrastructure, including supercomputers, cloud computing, and quantum computing, could also help facilitate modeling, simulation, and analysis. For example, the Air Force Research Laboratory has established a new supercomputer, named the Raider, which can calculate about 12 petaFLOPS, offering opportunities to run simulations at a higher level of accuracy and significantly accelerated timelines (Castrejon, 2023). Advanced computing capabilities will be essential to the joint test and mission engineering communities as M&S increasingly plays a major role in joint testing of complex missions.

Data and AI

The DoD set a vision for big data analytics, data governance, and AI in its DoD Data, Analytics, and Artificial Intelligence Adoption Strategy (DoD 2023). Providing data in a secure and trusted manner is critical to allowing AI and other digital engineering tools to function optimally and to enable reuse and analysis. Quality data are needed to build accurate models and insights. This need becomes more challenging with big data analytics and the collection, storage, and analysis of vast datasets to extract meaningful insights. Big data analytics can help joint test concepts identify trends to make better decisions and improve efficiency.

The effective use and application of data and AI will be foundational for joint test concepts to create a holistic picture of the joint environment and evaluate the mission scenarios within it. AI is increasingly being applied to various test and evaluation processes to enhance efficiency, decision-making and security. Some examples of this from the DoD's (2023) *Data, Analytics, and Artificial Intelligence Adoption Strategy* include

- **Cybersecurity Testing:** AI is used to simulate and detect cyber threats, vulnerabilities, and potential attacks on DoD networks. Automated tools help in identifying weaknesses and improving the overall cybersecurity posture.
- **Autonomous Systems Testing:** AI plays a crucial role in testing and evaluating autonomous systems, such as unmanned aerial vehicles and ground vehicles. It assists these systems to meet performance standards and can operate effectively in diverse environments.
- **Data Analysis for Intelligence:** AI applications are employed to analyze vast amounts of intelligence data, providing faster and more accurate insights. This aids in decision-making processes and enhances the efficiency of intelligence analysis.
- **Simulated Training Environments:** AI-driven simulations are used for training military personnel, creating realistic scenarios for test and evaluation of decision-making skills, strategic planning, and tactical execution.

These examples demonstrate how AI applications are strategically implemented within the DoD to address a range of current challenges, from cybersecurity to simulated training environments. The goal is to leverage technology to enhance capabilities, readiness, and overall effectiveness of joint tests. Advancements in AI and machine learning bring forth innovative opportunities and streamlined joint test concepts to help automate and optimize various evaluation tasks and processes, from automated testing and test generation to data collection, analysis, reporting, and more. When combined, these efforts have far-reaching immediate and future implications for joint testing that will enable the community to better validate and deliver the necessary critical technologies and systems to the warfighter to support a continued tactical, operational, and strategic military advantage. Adopting data, analytics, and AI technologies will help the DoD make decisions more accurately, efficiently, and expeditiously to support joint test and mission engineering (DoD Responsible AI Working Council, 2022).

TTPs in the Context of Mission Engineering

As TTPs evolve to reflect emerging technologies and warfighting needs, an opportunity exists to leverage data collected in operationally relevant testing to inform mission engineering and vice versa. DoD's Joint Test and Evaluation (JT&E) program considers emerging technologies and the increasingly complex and dynamic, joint, multi-domain operational environment to plan and execute test projects intended to deliver data-driven TTPs, concepts of operation, and other non-materiel solutions. Given the increased integration and dependencies of platform, network, and command and control solutions across the domains, JT&E's mission and its unique focus on system-of-systems testing is becoming increasingly critical to the department's strategic objectives. JT&E's extensive use of operational testing techniques and reach-back are essential to the adequate evaluation of the effectiveness of proposed solutions needed in operational plans across the combatant commands.

JT&E and mission engineering complement each other, especially when mission engineering is evaluating new capabilities for which the TTPs are still evolving. Choosing the specific scenarios and vignettes for test and analysis becomes increasingly difficult as the range of options grow. By working together, the JT&E program can integrate information from mission engineering, exercises, operational tests, and current operations to determine the best TTPs in a high-end fight. Understanding how to adequately represent an operationally realistic contested environment is essential to correctly develop and evaluate those TTPs. The use of JT&E joined with mission engineering will provide a powerful means of developing new, optimal joint warfighting TTPs to suffocate the enemy's ability to sustain the war.

Conclusion

The shift towards multi-domain operations and combined, joint all-domain command and control is driving a need for data-backed mission engineering. Underpinning the connections between mission engineering and operational and live fire test and evaluation, including JT&E, is the collection of operationally relevant data. The ultimate success of mission engineering depends upon the integration of extant and future operational test data to accelerate learning and increase the cycles of innovation. Conversely, the success of the transformation of operational and live fire test and evaluation depends upon the integration of mission engineering architectures to advance from a single system focus to a future joint warfighting concept and capability focus.

There are a variety of ways these two disciplines can enhance each other, but those opportunities will not be realized until operational testing becomes better integrated in system development. Overall, the emphasis needs to continue moving beyond dedicated operational testing to support production decisions, to gathering and integrating operationally relevant data to learn about systems' capabilities and how those capabilities support mission outcomes. Industry has embraced getting feedback from operations to improve its systems (to include hardware-based systems), using methods such as development, security, and operations (DevSecOps) and digital twins as part of Industry 4.0 (Madni et al., 2019). Operational feedback guides development of new features at Tesla, Amazon, and Netflix, helping these companies achieve a dominant position in their respective fields.

DoD's (2023) *Data, Analytics and Artificial Intelligence Adoption Strategy* has a similar focus on speed of delivery and continuous improvement, calling for "a tight feedback loop between technology developers and users through a continuous cycle of iteration, innovation and improvement of solutions that enable decision advantage." This strategy calls for the creation of effective, iterative feedback loops between developers, users, subject matter experts, and test and evaluation experts to ensure that the developed capabilities are more stable, secure, ethical, and trustworthy (see Figure 4).

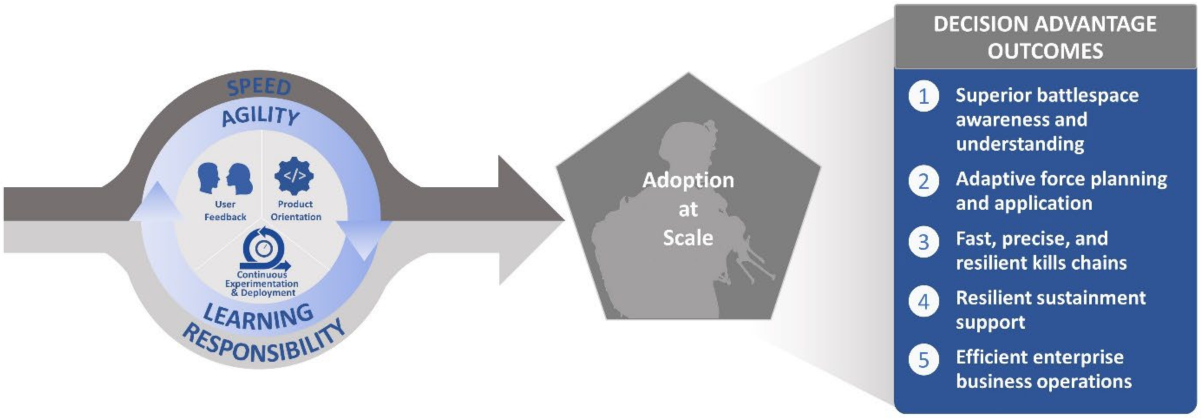


Figure 4. Agile Approach to Accelerate Decision Advantage.

Finally, as we look toward the future, the flexibility afforded by digital engineering and AI-enabled and autonomous capabilities, it is imperative to learn the most effective way to employ these capabilities. To accelerate the cycles of innovation, organizations need to agree on the key information needed from these learning efforts. The operational test community is working to quantify the benefits of digital engineering for operational testing and how improvements in knowledge management can be used to integrate all credible information in its evaluations. Models linking system designs and capability to mission outcomes have immense power to inform decision-making at multiple levels. Connecting mission engineering initiatives to operational and live fire test and evaluation, and JT&E, offers a tremendous potential to improve the ability to learn as an enterprise and effectively translate that learning to action, ensuring enduring mission success.

References

Air Force. (2023, March 20). *JB Elmendorf-Richardson selected to host new Joint Integrated Test and Training Center*. https://www.af.mil/News/Article-Display/Article/3334272/jb-elmendorf-richardson-selected-to-host-new-joint-integrated-test-and-training/

Albon, C. (2022, June 15). *Space Force envisions digital future for testing and training*. C4ISRNET. https://www.c4isrnet.com/battlefield-tech/space/2022/06/15/space-force-envisions-digital-future-for-testing-and-training/

Brown et al. (2023). Mission engineering for Hybrid Force 2025. *Naval Engineers Journal*, *135*(1), 141–153. https://www.ingentaconnect.com/contentone/asne/nej/2023/00000135/00000001/art00024

Castrejon, A. M. (2023, September 11). *AFRL's newest supercomputer 'Raider' promises to compute years' worth of data in days, saving time, money*. Air Force Research Laboratory Public Affairs. https://www.afrl.af.mil/News/Article-Display/Article/3521947/afrls-newest-supercomputer-raider-promises-to-compute-years-worth-of-data-in-da/

DoD. (2023, June 27). *Data, analytics, and artificial intelligence adoption strategy: Accelerating decision advantage*. https://media.defense.gov/2023/Nov/02/2003333300/-1/-1/1/DOD_DATA_ANALYTICS_AI_ADOPTION_STRATEGY.PDF

DoD Responsible AI Working Council. (2022, June 21). *Responsible artificial intelligence strategy and implementation pathway*. DoD, Chief Digital and Artificial Intelligence Office. https://www.ai.mil/docs/RAI_Strategy_and_Implementation_Pathway_6-21-22.pdf

Grady, C. W. (2023). Sharpening our competitive edge: Honing our warfighting capabilities through the joint warfighting concept. *Joint Force Quarterly*, *111*. https://ndupress.ndu.edu/JFQ/Joint-Force-Quarterly-111/Article/Article/3569518/sharpening-our-competitive-edge-honing-our-warfighting-capabilities-through-the/

Madni, A. M., Madni, C. C., & Lucero, S. D. (2019, January 30). Leveraging digital twin technology in model-based systems engineering. *Systems*, *7*(1). https://www.mdpi.com/2079-8954/7/1/7

Marler et al. (2022, June). *What is JADC2, and how does it relate to training?* RAND Corporation. https://www.rand.org/pubs/perspectives/PEA985-1.html

Office of the Under Secretary of Defense for Research and Engineering. (2023, October 1). *Mission engineering guide: Version 2.0*. Department of Defense. https://ac.cto.mil/wp-content/uploads/2023/11/MEG_2_Oct2023.pdf

Space Force. (2022, October). *Vision for: The national space test and training complex*. https://www.starcom.spaceforce.mil/Portals/2/NSTTC Vision_PA Final_1.pdf

Advantages of Using Complex Decision Support Tools in Planning Multi-Modal Test Programs

Milo Taylor—earned a Master of Science degree with the thesis option at Iowa State University where he studied applied mathematics. His thesis research involved implementing numerical methods for solving relativistic fluid problems. He is a modeling and simulation analyst with nearly a decade of experience in radar, autonomy, and electronic warfare. Taylor is currently a research scientist at the Georgia Tech Research Institute (GTRI) in the Electro-Optical Sensor Laboratory (EOSL). His recent work interests include Space EW and cislunar operations. [Milo.taylor@gtri.gatech.edu]

Craig Arndt—currently serves as a principal research engineer on the research faculty of the George Tech Research Institute (GTRI) in the System Engineering Research division of the Electronic Systems Lab. Arndt is a licensed Professional Engineer (PE), and has more than 40 years of professional engineering and leadership experience. Arndt holds engineering degrees in electrical engineering, systems engineering, and human factors engineering and a Master of Arts in strategic studies form the US Naval War college. He served as Professor and Chair of the engineering department at the Defense Acquisition University, and as technical director of the Homeland security FFRDC at the MITRE Corporation. In industry he has been an engineering manager, director, vice president, and CTO of several major defense companies. He is also a retired naval officer [Craig.arndt@gtri.gatech.edu]

David Zurn— received his bachelor's and Master of Science degrees in Electrical Engineering from the Georgia Institute of Technology in 1985 and 1990 respectively. Since joining GTRI's Electronic Systems Laboratory (ELSYS) in 2003, he has worked on a variety of EW-related research efforts including Radar Warning Receiver hardware and software development and test, Missile Warning System hardware and software test, and development of Hardware in the Loop (HITL) test solutions tailored to EW applications. Zurn is currently serving as the Division Chief of the Test Engineering Division within ELSYS. Zurn is a lecturer for the RWR Design short course offered through GT's Professional Education program. His recent research interest areas are Cognitive EW T&E and Space EW T&E. [David.zurn@gtri.gatech.edu]

Abstract

Emerging systems being tested in complex environments require the development of alternate test modalities, including hardware in the loop (HITL) and modeling and simulation (M&S) environments. The investment in these modalities are often significant. For example, testing the survivability of space system uplinks requires difficult over the air (OTA) testing or the development of threat models, orbital models, and propagation models tied together in a HITL or M&S testbed accurately simulating the problem. If properly designed, these testbeds could meet developmental or operational test requirements and potentially be used across a range of space acquisition programs. This research highlights challenges and approaches for developing alternate test modalities and proposes a multi-modal decision support tool for understanding the usage of the testbeds and evaluating tradeoffs. A specific example is explored for the space EW test use case.

Key words: Hardware in the Loop (HITL), Modeling and Simulation (M&S), Decision Support Tools

Executive Summary

This research investigates the challenges associated with evaluating the suitability of alternate test modalities for testing complex systems. We address the difficulty of testing complex systems and recognize that most complex systems are tested in operational test environments, which are referred to as over the air (OTA). Alternate test modalities (HITL and M&S) should also be considered for broader usage. However, it is not always clear where alternate modalities can be used and how much advantage can be achieved with alternate modalities.

A process and decision support tool would be effective for test planners in addressing these uncertainties. A framework is presented for the multi-mode test tool, along with an examination of a space uplink survivability example that is used with a first order implementation of the tool. The example allows us to identify several key uses for the tool:

- Evaluating trades between test objectives (quality, coverage, difficulty) and test modalities
- Understanding test use case to test modality mapping
- Test resource planning—understanding benefits and usages of alternate modalities.

The tool is discussed in the broader context of the system engineering process and common decision support tools. Finally, challenges and a potential way-forward with a tool of this type are discussed.

Background

There are a wide range of methods for testing complex systems throughout their life cycle. The selection of different methods to use at different times has been developed over time and is now incorporated into policies and procedures at different acquisition and test organizations. Most of these practices were established well before the advent of digital engineering processes and do not take into consideration the capabilities of alternate test modalities, including hardware in the loop (HITL) and modeling and simulation (M&S), incorporating digital twins and other digital representations of the system under test.

The introduction of digital engineering has changed both the methods we used to develop defense systems and the timelines associated with the development of those systems. The reduction in the time it takes the Department of Defense (DoD) to develop, test, and field a system is a high priority for the DoD acquisition leadership.

Traditionally, test and verification of a defense system is a complex and time-consuming enterprise. As defense systems continue to grow in complexity, the resources needed to test and verify these systems also grow. There is however a need to reduce, not grow the timelines for testing.

There are a number of different ways that systems can be tested. These different test methods or modes offer the opportunity to verify system performance in different ways. However, the test modes are significantly different. The principle modes are live range testing (referred to as over the air (OTA) testing for this report), HITL, and M&S. Each of these modes have advantages and disadvantages. Moreover, these different test modes can be conducted at different times in the life cycle of the system: development, manufacturing, and operation.

The need to reduce testing time and to accelerate system development and fielding favors early testing using HITL and M&S tools where possible.

The test planner faces challenges in determining the suitability of different modes of testing at different times in the life cycle of the system. Tools and processes are required to help the test planner understand the tradeoffs involved in using alternate modes and the test cases for which each mode is most suitable. If used correctly, these tools can inform test plans that meet program cost and schedule while maintaining high confidence in the performance of these systems.

Test Modalities

Over the Air

Over the air (OTA) testing is a traditional testing mode in which the system under test is placed in a real-world environment. For example, at the Redstone Test Center, the Open-Air Range provides field testing for sensor and seeker systems. Such test ranges allow for the characterization of system targets and environments the system may operate in. System components may also be tested in a controlled environment. An environmental chamber at such a range further provides a real-world simulation of conditions the system under test may face (U.S. Army, 2024). In the context of space vehicles, the live test range may be the actual orbit of the satellite. Uses of OTA for in-orbit testing include future efforts in upgrading sensors on existing satellites whose primary function is space domain awareness (Albon, 2024).

Due to the nature of OTA, the clearest advantage is the level of fidelity afforded by implementing real systems and effectors.

However, considerations for OTA testing stem mostly from its level of fidelity and the fact that the system under test and other test resources exist in the real world. For example, one must consider emissions control (EMCON) when dealing with multiple electromagnetic signals that may interfere or be exposed to unauthorized monitoring.

Hardware in the Loop

Hardware in the loop (HITL) testing is a T&E solution that provides a blend of real-world components and simulation facilities. For example, for nearly 30 years, the U.S. Army Redstone Test Center (RTC) has served as a U.S. Army Test and Evaluation Command to provide T&E support and facilities for various customers. In terms of HITL, the RTC provides T&E for missile seekers by combining traditional T&E with virtualized HITL and M&S environments (U.S. Army, 2024).

An example of HITL in satellite testing is known as a FlatSat. A FlatSat is a "high fidelity electrical and functional representation" of the satellite bus (Amason, 2008). For NASA's Solar Dynamics Observatory (SDO), it is a test bed for integration and test, flight software, and flight operations.

Some benefits of HITL testing include: ability to perform repetitive tests, non-destructive tests where applicable, and closed-loop testing to minimize external factors. Some benefits of a FlatSat in particular include pre-launch flight software development and verification due to the fact that a physical representation is being tested on the ground in a lab.

Modeling and Simulation

Digital modeling and simulation (M&S) testing is the means of using digital models of the system, its processes, and the environment to test system performance. Digital M&S testing is a mode in which the tests are fully implemented digitally. However, the models of the systems involved and environmental factors are driven by data and intelligence.

Each component of a digital model—whether the system itself or the environment it will operate in—may vary in its scope or fidelity. For example, in the context of DoD applications, M&S software suites may be primarily suited for different fidelities such as mission-level modeling in the Advanced Framework for Simulation, Integration and Modeling (AFSIM) model (*AFSIM, AFRL*, 2024) or at the campaign level as in the Synthetic Theater Operations Research Model (STORM; *STORM, AFRL*, 2023). However, with growing computational capabilities and years of software development, some of these tools may allow the user to operate with others at varied levels of fidelity, or span these levels themselves.

Some advantages of digital M&S testing include the relatively low cost to represent the system and the ability to test under various levels of complexity. Some sources of difficulty can be alleviated if there is a standardization of M&S principles and practices. The recent increased DoD adoption of digital twins for M&S has addressed some of these difficulties and is accelerating usage of M&S for test.

In 2023, the Office of the Director, Operational Test and Evaluation (DOT&E) Strategy Implementation Plan laid out five strategic pillars. In particular, the plan raises Pillar 4, "Pioneer T&E of weapon systems built to change over time" which focuses on standardizing and promoting the use of digital tools. This plan specifically calls for increased usage of digital twins as well as other tools in digital engineering (Director, Operational Test and Evaluation, 2023c).

According to the Digital Twin Consortium, a digital twin is "a virtual representation of real-world entities and processes, synchronized at a specified frequency and fidelity" (Digital Twin Consortium, 2020). Digital twins may come in a myriad of different forms, but an important feature is the synchronization with its real-world counterpart over the system's life cycle.

General Framework for Planning Multi-Modal Tests

We've defined the need for efficient testing as systems grow more complex and acknowledged that modern systems require a mix of test modalities, which we've defined as OTA, HITL, and M&S. Each modality brings with it advantages and disadvantages typically in the form of fidelity, coverage, and cost. The challenge for the test planner is to determine which individual test use cases are best-suited to specific modalities. The current process for developing test plans looks at available test resources and through interaction with experts, determines the best approach for planning individual test use cases and determining the best modality for each use case. This becomes both difficult and inefficient however, as the size and complexity of systems grow. In the following section we propose a general approach for aiding the tester in selecting the "best" test mode for specific test use cases.

The design of any test is a tradeoff between competing objectives such as:

- Quality
  - Fidelity—the level of detail which the test replicates in the operational environment that the system will be operating in.
  - Repeatability—the ability to produce the same results if the test is conducted multiple times with no change in parameters
  - Reliability/Confidence—the measure of how well results represent the real world and the sensitivity to external factors—a function of the number of test data points collected (this is determined as experimental design)
- Coverage—the part or percentage of the system performance envelope that the test verifies.
- Difficulty
  - Cost—a measure of the affordability of a test
  - Schedule—a measure of the timeliness of a test
  - Risk—an assessment of whether the test will function as intended and provide usable data

An ideal test program will experience high quality, extensive coverage, with low difficulty. This is not achievable because some factors improve at the expense of others. The fundamental tradeoffs most designers encounter due to test resource limitations are:

- As quality increases, coverage generally decreases
- As quality increases, difficulty increases
- As coverage increases, difficulty increases

Trade-offs are always present in a given test design, particularly for the specific test mode chosen. Moreover, we can generally say that OTA, HITL, M&S test modes typically come with the following objective tradeoffs:

1. OTA—quality high, coverage low, difficulty high
2. HITL—moderate quality, moderate coverage, moderate level of difficulty
3. M&S—lowest quality, highest coverage, lower difficulty.

Test resource planners should understand these objective trade-offs to determine the "best" mode for each use case. The following framework is proposed to aid in this determination.

Define a Use Case set, composed of a set of Test Categories:

USE_CASE (i,j,k) = {FUNC(i), ENV(j), ENG(k)}

Where Test Categories are defined as:

CAT = {FUNC, ENV, ENG} with

FUNC = {$func_1$, $func_2$, … $func_x$}, a set of Functions or Functional modes,
ENV = {env_1, env_2, … env_y}, a set of Environmental variants,
ENG = {eng_1, eng_2, … eng_z}, a set of Engagement variants.

Next define a set of Test Modes over which to evaluate the Use cases:

MODE = {OTA, HITL, M&S}.

Finally, define a Test Objective set that supports the evaluation of use cases. The Test Objective set is defined as:

OBJ = {Quality, Coverage, Difficulty}.

Note that we've simplified Test Objectives for this general framework. In reality, the Quality and Difficulty objectives should be decomposed into the components described above, scored, and combined to provide overall Quality and Difficulty scores.

To begin the evaluation, each combination of Test Category (CAT) and Mode (MODE) are scored for each Objective (OBJ). As an example, a score would be assigned to func1, using the OTA Mode, for the Quality test objective. This scoring would be performed for all members of the FUNC, MODE, OBJ sets, ENV, MODE, OBJ sets, and ENG, MODE, OBJ sets, creating a three-dimensional scoring array SCORE with CAT, MODE, and OBJ vectors. A representation of the array is shown below with Test Category and Mode shown for each Objective.

Quality Scores

		Mode OTA	Mode HITL	Mode M&S
CAT	func1	Q_{1O}	Q_{1H}	Q_{1M}
	.	Q_{2O}	Q_{2H}	Q_{2M}
	env1	.	.	.
	.	.	.	.
	eng1	.	.	.
	.	Q_{nO}	Q_{nH}	Q_{nM}

Coverage Scores

		Mode OTA	Mode HITL	Mode M&S
CAT	func1	C_{1O}	C_{1H}	C_{1M}
	.	C_{2O}	C_{2H}	C_{2M}
	env1	.	.	.
	.	.	.	.
	eng1	.	.	.
	.	C_{nO}	C_{nH}	C_{nM}

Difficulty Scores

		Mode OTA	Mode HITL	Mode M&S
CAT	func1	D_{1O}	D_{1H}	D_{1M}
	.	D_{2O}	D_{2H}	D_{2M}
	env1	.	.	.
	.	.	.	.
	eng1	.	.	.
	.	D_{nO}	D_{nH}	D_{nM}

Figure 2. Scoring Array

Next a set of Use cases, combining the FUNC, ENV, and ENG Test categories is created. The total number of Use cases depends on the number of combinations of the variants in each Test category. A simple example might be a system with four discrete Functional modes (FUNC), three Environmental variants (ENV), and two Engagement (ENG) variants. In this case the system would have a maximum total of 512 (2^9) total potential use cases. Some combination of variants might not make sense so this is the maximum number of Use cases. For each use case, a score is calculated for each objective. The score combines the Scoring array entries for the Test Categories included in that Use case. Scores are calculated as follows:

Quality_Score (Use Case, Mode) = SCORE (FUNC (Use case), Mode, Quality) + SCORE (ENV (Use case), Mode, Quality) + SCORE (ENG (Use case), Mode, Quality)

Coverage Score (Use Case, Mode) = SCORE (FUNC (Use case), Mode, Coverage) + SCORE (ENV(Use case), Mode, Coverage) + SCORE (ENG(Use case), Mode, Coverage)

Difficulty_Score (Use Case, Mode) = SCORE (FUNC (Use case), Mode, Difficulty) + SCORE (ENV (Use case), Mode, Difficulty) + SCORE (ENG (Use case), Mode, Difficulty)

Finally, a Utility score is calculated for each Mode in each Use case. The Utility score is a weighted sum of the objective scores defined above.

Utility (Use Case, Mode) = Quality_Score (Use Case, Mode)*w1 + Coverage_Score (Use Case, Mode)*w2 + Difficulty_Score (Use Case, Mode)*w3

The w1, w2, w3 weight values are determined based on the test type and system complexity. For example, training, developmental testing (DT), and operational testing (OT) each have increasing weight placed on quality with OT designated as highest required quality. Increasing system complexity might place higher weight on coverage.

For each use case, the Utility scores for each of the three Test modes are compared. The highest Utility score indicates the "optimal" test mode for that use case.

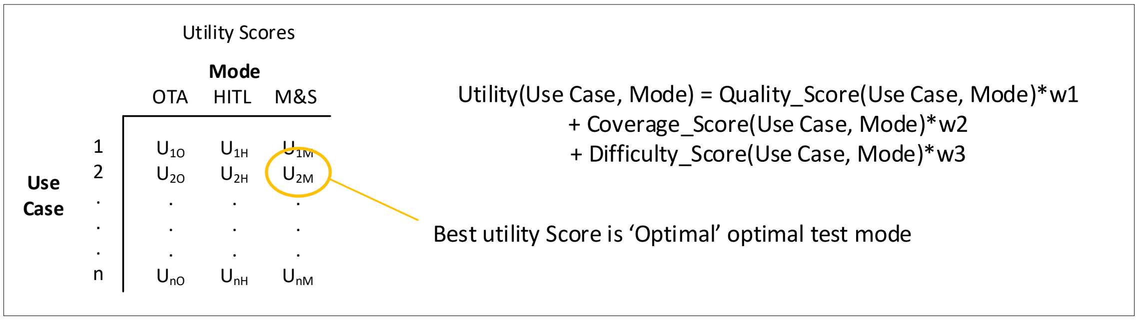


Figure 3. Total Utility Function

A simplified schematic of the process is shown below.

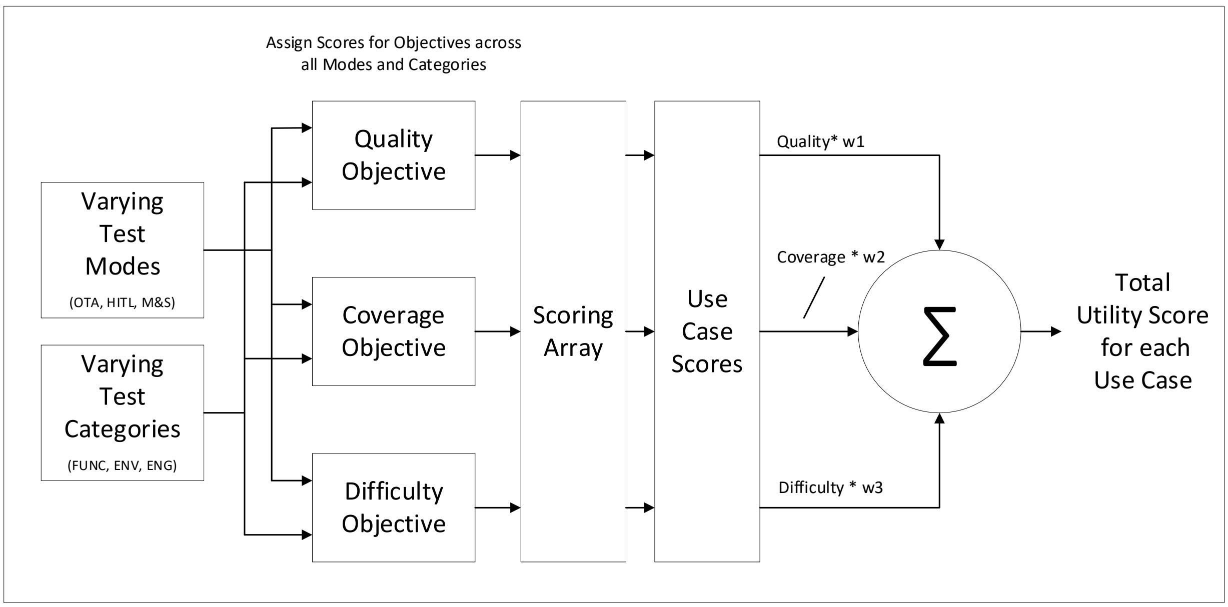


Figure 4. Simplified Utility Scoring Schematic

Multi-Modal Test Tool—Uplink Survivability Test Example

Counterspace threats come in various forms ranging from direct-ascent anti-satellite weapons to cyber-attacks. Electronic warfare (EW) poses a particularly unique set of challenges to successful operations in space. Indeed, the most modern militaries consider EW to be an essential facet of warfare, and have incorporated jamming and anti-jamming counterspace capabilities (Defense Intelligence Agency, 2022). Likewise, military powers have incorporated EW to secure navigational and informational superiority. In offensive electronic warfare, the objective is to disrupt, deny, degrade, destroy, or deceive communications or target acquisition.

At a basic level, a satellite communications (SATCOM) set up is composed of three segments (NASA, 2024):

1. Space segment: a collection of space vehicles
2. Link segment: the functional segment consisting of signals from the ground (uplink), transmitting data down to the ground (downlink), and transmitting and receiving data to and from other satellites (crosslink)

3. Ground segment: assets located on the ground (or in sometimes air, land, and sea) such as ground control or user terminals

The ground segment may be decomposed further into the control terminals, user terminals, and infrastructure.

An adversarial actor may interfere with SATCOM by introducing jamming or spoofing. Uplink survivability (ULS) testing determines how well the satellite under test (SUT) performs in the presence of such jamming of the uplink signals. In uplink jamming, a threat system specifically interferes with a signal originating from the ground segment meant for the space segment. The purpose is to deny or degrade the reception at the satellite receiver in order to prevent communication, increase error rates, or decrease throughput.

ULS Test Description

The SUT is a blue satellite in orbit (any orbit type) capable of transmitting and receiving either a data link or telemetry, tracking, and command (TT&C). The emphasis is on testing and evaluating the system's operational capability in the presence of a corrupted uplink signal.

On the blue side, the ULS test case is comprised of the SUT (a representation of the space vehicle, or the vehicle itself), a ground control terminal, relevant infrastructure depending on the test modality, test control instrumentation, and possibly a user downlink terminal.

On the threat side, the ULS test case is comprised of at minimum a simulator for the threat jammer. The threat jammer may employ either basic noise jamming, or more advanced techniques. The uplink survivability test case may also be generalized to consider the survivability of a constellation of satellites against multiple sources of interference for an M v N engagement.

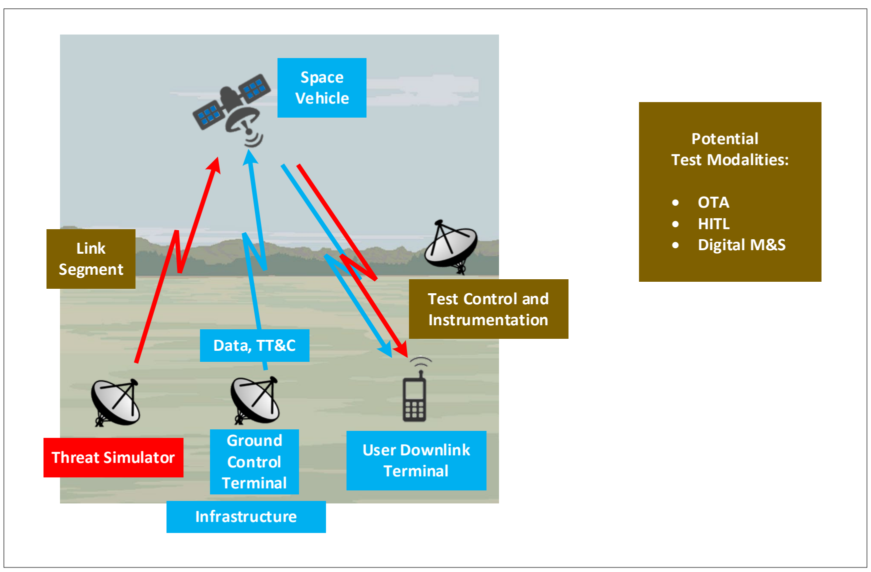


Figure 5. Basic ULS Test Components

Framework Applied to ULS test

Next, the general framework for planning multi-modal tests defined above will be applied to the ULS test problem. Recall, the framework is intended to identify optimal test modes for each test use case.

The first step is to establish use cases for the ULS test example. A use case is a basic test case composed of categories that reflect a specific operating condition of the test article. An example for ULS might be testing a data link in the presence of basic interference, in a clear environment, with a single interference source. We can designate the following categories for ULS testing:

- Function—Data link, TT&C link
- Interference Type—Basic, Advanced
- Environment—Clear, Obscured
- Engagement—1v1, M v N

Note that we've added Interference Type to the three test categories defined above (Function, Environment, Engagement). Interference could have been added as an additional sub-category of environment. We chose to break it out into its own category because it's a key component in ULS testing. Note also that this is a simplified example—a real test design would incorporate a wider variety of types for each category.

Next, test objectives are established. For this example, we'll use Quality, Coverage, and Difficulty as first order test objectives. As noted above, formal detailed analysis would decompose these into sub-objectives (for example, Quality would be decomposed into Fidelity, etc.) for a more accurate assessment.

The method described above is used for scoring each objective based on the Function, Interference Type, Environment, and Engagement categories. The scoring is done for the three test modes, for each category. The actual scoring should be done by test designers with knowledge in the test domain and knowledge in the three test modes. For this exercise, we scored by assigning numeric values from 1–9. To simplify this initial analysis, the three modes were ranked with 7 assigned to the highest-ranking mode, 5 assigned to the next highest, and 3 assigned to the lowest mode. Scores of 1 or 9 were assigned to "edge cases" where an extreme score is justified.

	Use Case Category																							
	Link						Interference						Environment						Engagement					
	Data			Control			Basic			Advanced			Clear			Obscured			1v1			MvN		
Objective	**O**	**H**	**M**	**O**	**H**	**M**	**O**	**H**	**M**	**O**	**H**	**M**	**O**	**H**	**M**	**O**	**H**	**M**	**O**	**H**	**M**	**O**	**H**	**M**
Quality	9	5	3	9	5	3	9	5	3	9	7	3	7	5	3	3	7	3	7	5	3	1	5	5
Coverage	3	5	7	3	5	7	3	5	7	3	5	7	3	5	7	3	5	7	3	5	7	1	1	9
Difficulty	1	5	7	1	5	7	3	7	5	1	7	5	1	5	7	1	5	7	3	5	7	1	1	9
Scoring:	Rate 7,5,3 with 7 = best; allow extremes for edge cases (9,1)																							

Figure 6. Scoring Array for ULS Example

Next, intermediate scores are calculated for each use case, for Objective and Test mode, as shown in **Figure 16Figure 16** below. The intermediate score sums the Scoring array entries for the Test Categories included in that use case.

In the final step, we calculate Utility values for each use case by combining scores determined for each Category and Mode in the Scoring Array. The Total Utility is a simple weighted sum of Category scores for each Objective for each Mode as described in the framework description above. See Figure 6 for results using even weighting values (w1=w2=w3=5) for each objective.

Use Case					Intermediate Scores									Total Utility Scores		
					Quality			Coverage			Difficulty					
Use Case #	Function	Interference	Environment	Engagement	OTA	HITL	M&S	OTA	HITL	M&S	OTA	HITL	M&S	OTA	HITL	M&S
1	Data Link	Basic	Clear	1v1	32	20	12	12	20	28	8	22	26	260	310	330
2	Data Link	Advanced	Clear	1v1	32	22	12	12	20	28	6	22	26	250	320	330
3	Control Link	Basic	Clear	1v1	32	20	12	12	20	28	8	22	26	260	310	330
4	Control Link	Advanced	Clear	1v1	32	22	12	12	20	28	6	22	26	250	320	330
5	Data Link	Basic	Obscured	1v1	28	22	12	12	20	28	8	22	26	240	320	330
6	Data Link	Advanced	Obscured	1v1	28	24	12	12	20	28	6	22	26	230	330	330
7	Control Link	Basic	Obscured	1v1	28	22	12	12	20	28	8	22	26	240	320	330
8	Control Link	Advanced	Obscured	1v1	28	24	12	12	20	28	6	22	26	230	330	330
9	Data Link	Basic	Clear	MvN	26	20	14	10	16	30	6	18	28	210	270	360
10	Data Link	Advanced	Clear	MvN	26	22	14	10	16	30	4	18	28	200	280	360
11	Control Link	Basic	Clear	MvN	26	20	14	10	16	30	6	18	28	210	270	360
12	Control Link	Advanced	Clear	MvN	26	22	14	10	16	30	4	18	28	200	280	360
13	Data Link	Basic	Obscured	MvN	22	22	14	10	16	30	6	18	28	190	280	360
14	Data Link	Advanced	Obscured	MvN	22	24	14	10	16	30	4	18	28	180	290	360
15	Control Link	Basic	Obscured	MvN	22	22	14	10	16	30	6	18	28	190	280	360
16	Control Link	Advanced	Obscured	MvN	22	24	14	10	16	30	4	18	28	180	290	360

Test Objective	Weight
Quality	5
Coverage	5
Difficulty	5

Figure 7. Use Cases and Utility Values for ULS Example—Equal Objective Weighting

ULS Test Case Observations

Green highlighted cells indicate optimal Test Modes for each use case. Red highlighted cells indicate lowest total score for each use case. Note that in this example, all of the use cases selected M&S as the optimal mode. Closer examination of the intermediate scores show that the M&S modes scored consistently higher for Coverage and Difficulty. This, combined with the fact that all Objectives were weighted equally (at 5), drives consistently higher total scores for M&S.

Based on this first example we can see that total utility scores heavily depend on Objective weightings. To explore this further, a first order sensitivity analysis by weighting was conducted. This sensitivity analysis yields the following results for test modes with highest utility score:

1. **Evenly weighted**—M&S for nearly all use cases, with several HITL
2. **Heavy weighting towards Quality**—mix of OTA, HITL use cases
3. **Heavy weighting towards Coverage**—all M&S
4. **Heavy weighting towards Difficulty**—all M&S
5. **OT weighting** (high Quality, low Coverage, medium Difficulty)—mix of OTA, HITL, M&S
6. **DT weighting** (high Quality, high Coverage, medium Difficulty)—mix of OTA, HITL, M&S

Case		Weight			# Modes w/Highest Utility Score		
		Quality	Coverage	Difficulty	OTA	HITL	M&S
1	Evenly weighted	5	5	5	0	2	16
2	Quality weighted	8	2	2	8	8	0
3	Coverage weighted	2	8	2	0	0	16
4	Difficulty weighted	2	2	8	0	0	16
5	Operational Test	8	2	3	4	10	2
6	Developmental Test	7	4	3	2	6	8

Figure 8. Sensitivity Analysis Showing Variance in Optimal Mode Selection as a Function of Weighting

The sensitivity analysis underscores the basic Fidelity versus Coverage tradeoff inherent in most tests. We expect that if the Fidelity (Quality) is heavily weighted and coverage is rated low (Case 2 in table above), then OTA would be a preferred choice. If Coverage is heavily weighted (Case 3) then M&S is the preferred choice.

The results are shown in Figure 8 for the set of Operational Test weights, which reflect a more realistic weighting scheme.

Use Case					Intermediate Scores									Total Utility Scores		
					Quality			Coverage			Difficulty					
Use Case #	Function	Interference	Environment	Engagement	OTA	HITL	M&S	OTA	HITL	M&S	OTA	HITL	M&S	OTA	HITL	M&S
1	Data Link	Basic	Clear	1v1	32	20	12	12	20	28	8	22	26	304	266	230
2	Data Link	Advanced	Clear	1v1	32	22	12	12	20	28	6	22	26	298	282	230
3	Control Link	Basic	Clear	1v1	32	20	12	12	20	28	8	22	26	304	266	230
4	Control Link	Advanced	Clear	1v1	32	22	12	12	20	28	6	22	26	298	282	230
5	Data Link	Basic	Obscured	1v1	28	22	12	12	20	28	8	22	26	272	282	230
6	Data Link	Advanced	Obscured	1v1	28	24	12	12	20	28	6	22	26	266	298	230
7	Control Link	Basic	Obscured	1v1	28	22	12	12	20	28	8	22	26	272	282	230
8	Control Link	Advanced	Obscured	1v1	28	24	12	12	20	28	6	22	26	266	298	230
9	Data Link	Basic	Clear	MvN	26	20	14	10	16	30	6	18	28	246	246	256
10	Data Link	Advanced	Clear	MvN	26	22	14	10	16	30	4	18	28	240	262	256
11	Control Link	Basic	Clear	MvN	26	20	14	10	16	30	6	18	28	246	246	256
12	Control Link	Advanced	Clear	MvN	26	22	14	10	16	30	4	18	28	240	262	256
13	Data Link	Basic	Obscured	MvN	22	22	14	10	16	30	6	18	28	214	262	256
14	Data Link	Advanced	Obscured	MvN	22	24	14	10	16	30	4	18	28	208	278	256
15	Control Link	Basic	Obscured	MvN	22	22	14	10	16	30	6	18	28	214	262	256
16	Control Link	Advanced	Obscured	MvN	22	24	14	10	16	30	4	18	28	208	278	256

Test Objective	Weight
Quality	8
Coverage	2
Difficulty	3

Figure 9. Use Cases and Utility Values for ULS Example—OT Objective Weighting

Basic trends in the intermediate scores correspond with our understanding of tradeoffs associated with the basic test modes:

- For the Quality objective, OTA scored highest—this is expected given that OTA most closely resembles real-world conditions. Note that scores for OTA Quality decrease for obscured use cases and MvN use cases because these are more difficult to replicate in the OTA mode.
- For the Coverage objective, M&S scored highest for all uses cases. It is expected that properly constructed M&S environments should provide the best coverage.
- For the Difficulty objective, M&S scored highest for all use cases. It is assumed that once the M&S environment is set-up, the difficulty associated with running these tests is lowest for all test modes. Note that these scores assume that the M&S environment has been constructed, Blue and Red models developed, and the combined environment has been verified and validated. This is likely a faulty assumption for ULS M&S tools given the current state of M&S tool development in the EW domain.

The optimal test modes selected for test use cases make intuitive sense when considering each use case in detail:

- Use cases 1–4 selected OTA because Quality was scored highly because these use cases called for a clear environment.
- Use cases 5–8 selected HITL because these cases called for an obscured environment, which is difficult to achieve consistently in an OTA mode, but can be

more readily simulated in a HITL setup. This drove Quality scores for HITL higher and OTA lower.

- Use cases 9–16 selected HITL and M&S because these use cases require MvN engagement which disadvantages OTA scoring. The MvN use cases require multiple orbital SUTs and interference sources, which are difficult to achieve with an OTA test.

Multi-Modal Test Tool—General Observations

The multi-modal test tool we've presented is suitable for evaluating basic trades between Quality, Coverage, and Difficulty for different test modes and test use cases. It is not intended for detailed test planning such as assigning specific use cases to specific modes. More granularity is required for the test use case definition. This should be done by carefully examining the basic test scenario and defining a wider range of test categories, in much greater detail. Additionally, the Objective functions need to be refined. As presented, "Quality" encompasses a variety of components (Fidelity, Repeatability, Confidence/Reliability) as does Difficulty. These need to be split into distinct attributes that can be scored separately, then combined into the appropriate objective. Note however that the number of objectives used to calculate the total utility function needs to remain small. Adding additional, non-critical objectives to the utility function will dilute the effect of critical objectives in expressing utility.

A more fully developed multi-modal test tool could be used by test planners to:

- Understand the key modalities required for a test campaign, which would drive test planning and near-term test resource development
- Assign specific use cases to specific test modes/resources as part of test planning
- Understand the impact of emphasizing one test objective over another (i.e., Quality versus Coverage, etc.)

It is essential to recognize, however, that the tool is entirely dependent on subject matter experts (SMEs) providing accurate scoring. Fundamentally the tool relies on the SMEs to score individual categories versus objective functions. It does not capture the actual relationship between a category and objective function. Indeed, subtle tradeoffs such as fidelity versus coverage are captured through SME scoring, not through tool design.

Given the critical nature of scoring in achieving reliable results, users of the tool need to pay close attention to scoring methods. SMEs should be carefully chosen, and should independently assign scores which are then compared for variance. If the variance between SME scoring is significant then the scoring should be re-evaluated. Other methods for validating scoring should be developed as well.

The tool relies on the user assigning appropriate weighting to the Quality, Coverage, and Difficulty objectives. The utility function is heavily dependent on weighting. In fact, a biased user of the tool can achieve whatever result is desired by arbitrarily adjusting weights. It is recommended that prior to scoring, careful consideration be given to establish weighting values appropriate to the test application. Application of this tool for an operational test may prioritize Quality over Coverage and Difficulty, but a test designed for science and technology (S&T) application may favor lower Difficulty at the expense of Quality.

The multi-modal test tool could also be used for long term test resource development. Consider an acquisition organization developing a new capability and trying to determine test resources required. The tool could be used to understand the Quality, Coverage, and Difficulty trades associated with the resource test modes. This could drive resource planners to develop specific solutions in key modes as indicated by the tool. It has been observed that test resource developers generally support investment in OTA resources but are reluctant to invest in HITL and M&S resources because it is difficult to see the value provided by these alternate modes. This tool illuminates the trade-space for the three test modes, exposing benefit for alternate modes. Additionally, the tool may help more directly with resource acquisition—understanding use cases and test mode mapping can aid in developing capabilities, needs, and requirements for the test environments associated with these modes.

To facilitate its use for resource development, it may make sense to develop a "Development Difficulty" objective associated with test resource development. The Difficulty objective described above relates to difficulty in actually performing the test and not in developing a resource. Introducing the Development Difficulty objective ensures that the utility of each mode also reflects development difficulty. This may not be an issue for existing test resources, and if that's the case, the user can weight Development Difficulty at zero. If, however, a resource does not exist or significant effort is required to develop components of a resource, the overall utility of the resource is negatively impacted. For example, using M&S may provide great coverage with little relative difficulty in running the test. Yet if significant effort and risk is involved in developing and validating Blue and Red models, the M&S mode may not in reality provide much utility. Indeed, the user of the tool can be misled about the utility of various modes if they are assumed to provide benefit but end up requiring substantial resources to develop.

Multi-Modal Test Tool in the System Engineering Context

The systems engineering process is critical to all aspects of the development and testing of DoD systems. Within this process are a number of key steps which include requirements development, design, and test. Testing and verification are critical to the system engineering process because the testing and evaluation are needed in the design and development process and also needed to verify the performance of the end product and the manufacturing of the system before deployment. As a result, there are a wide range of tests that are conducted throughout the system life cycle. Testing and evaluation can be done in many different ways and times. To better understand the scope of different means of testing, we can look at testing along several different dimensions. First look at the purposes for testing. Testing can be performed to a) determine subsystem performance in design, b) determine overall system performance in design, and c) verify performance of the system for operational suitability, survivability, and effectiveness. Second, different actors conduct the testing, including designers, manufacturers, and different government organizations (including users). Third, testing is conducted at different times, including before the program starts (for legacy system parts), during development, during formal system verification, and after the system is in operation. The multi-modal test tool should be considered for test planning for all of these test types.

The key element of the defenses systems engineering process are captured in the systems engineering "V," as shown in Figure 9.

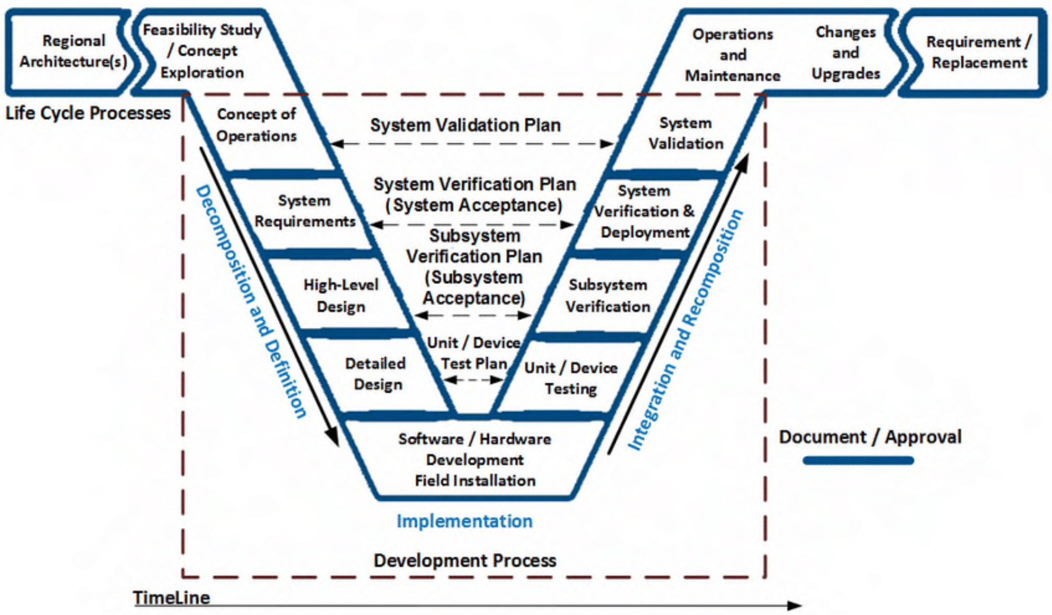


Figure 10. System Engineering "V"

As we can see from the systems engineering "V," the test process needs to be engaged throughout the life cycle. Although most tests and other verification activities are conducted after the system has been designed and built, the planning for test and verification should be an integral part of the development of system requirements and all aspects of the design and realization process. This leads to opportunities to do different tests very early in the development process and different modeling and simulation opportunities.

Decision Support Tools

The multi-modal test tool is a simple form of a more complex decision support tool. It may be beneficial to apply lessons learned from existing decision support tools as the multi-modal tool is more fully developed. Some basic background relating to decision support tools is presented below.

Decision support tools are used across a wide range of different domains to help analyze different courses of action. Over the past few years decision support tools have advanced significantly. The most common technology that has emerged is multi-dimensional decision frontiers. This mathematical analysis allows the user to evaluate complex multi-dimensional trade spaces. Trade space analyses are needed to support key decision makers, and some questions critical to informing these decisions are not well-addressed via traditional, more globally focused analyses. Systems engineering questions unique to a given system or problem will often require similarly unique analytical workflows supported by contextually relevant data. Multiple specific systems engineering insights can be gleaned from exploration of specific analysis pathways rather than over-simplified global analysis. To address this issue, the Georgia Tech Research Institute (GTRI) has sought to tie analytical components (building blocks such as sensitivity analyses, regression models, etc.) to data pipelines relevant to the question we are trying to ask. The question in the case of optimizing different combinations of testing methods and testing times is what combination of test parameters will maximize the quality of the test while minimizing cost and schedule.

As it is used here, context can refer to how the additive value of a system varies between stakeholders or temporal differences in a system's application over its life cycle that impact its perceived usefulness. A major understanding from GTRI's efforts is that generating a trade space from various models is not a trivial task if the goals are to achieve flexibility, scalability (often via properly orchestrated modularity), and efficiency of the process. Also, a use case has a specific path through a networked workflow. In addition, these goal characteristics are often defined according to life cycle stage or blur across several stages—care must be taken to operationalize appropriately. Specifying the precise way in which any analytical construct applies to trade space analysis and also its specific life cycle context is critical to future synthesis with other methods. Composability and traceability of constructs is key to future maturation using other methods in tandem. GTRI discovered through this work that the degree of modularity and the extent of the abstract description necessary to define the problem in a way that is directly executable is strongly linked and tremendously important to usability by a person and reusability in a computational environment. An example decision support tool is included in Figure 10 (Ender, 2014).

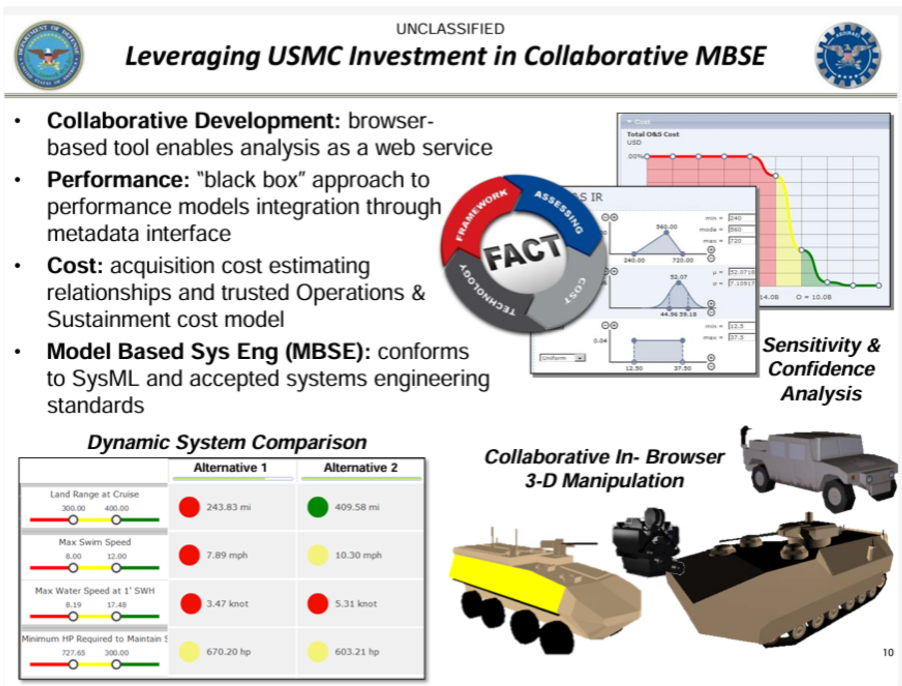


Figure 11. Example Decision Support Tool

Challenges

Several challenges must be addressed to more fully develop the multi-mode test tool.

The example tool developed for this research and presented above is specific to the ULS test application. How readily can the tool be generalized for other test applications? Ideally a fully developed tool could be re-used for different test applications without major re-work. It appears that the framework, process, and objective functions are generizable. However, the specific test use cases and their component categories are unique to a given test application and any future version of this tool must allow the user to specify use cases and categories in an efficient way. Interestingly, the basic OTA, HITL, and M&S test modes as defined may vary depending on system complexity and type. For this research we've implicitly defined the test modes according to their degree of virtualization:

- OTA—comprises no virtualization (as close to real world as possible)
- HITL—comprises no virtualization of the test article, with other elements (environment, interference) virtualized
- M&S—virtualizes all components

In reality, as a system gets more complex, the OTA-HITL-M&S distinctions may become a continuum of modalities where various components of the system or parts of the test scenario (aspects of the environment for example) are virtualized in a way that makes sense for efficient testing. The tool may need to be designed to account for this.

Another critical challenge is performing verification and validation (V&V) for the decision support tool. Verification (establishing that the tool is performing the way it's been designed) should be straightforward, using synthetic data. Validation on the other hand needs to establish that the tool is meeting the needs of the test planner. This involves determining whether the methods and processes underlying the tool are providing useful predictions. Ideally, validation would compare the tool test mode recommendations to historical data, but specific use case/test mode mapping data may be limited for specific test applications. Alternatively, independent review of tool output across a variety of test applications may be required for validation.

Conclusion

A multi-modal test decision support tool could be effective for aiding test planning and test resource development planning for complex systems. A practical framework has been created for a multi-mode test tool, which if developed into a formal decision support tool, could be used to:

- Evaluate trades between test objectives (Quality, Coverage, Difficulty) and OTA, HITL, and M&S test modalities.
- Understand the test use case to test modality mapping.
- Aid in test resource planning by highlighting benefits and usages of alternate modalities.

The authors recommend continuing this research by fully developing the ULS test tool presented here. There is an opportunity to collaborate with test planners and test resource acquisition professionals who are currently engaged in determining ULS test methods and doing specific ULS test planning. These SMEs could provide inputs for tool development and scoring. Ideally the tool outputs would help to inform their decision making and lead to more efficient testing. The team should also leverage existing decision support tool research, identifying well-developed frameworks and interfaces, enabling the efficient development of a multi-mode test tool.

References

AFSIM, AFRL. (2024). https://www.wpafb.af.mil/News/Art/igphoto/2001709929/

Albon, C. (2024, March 27). *Space Force to upgrade sensors for in-orbit testing, training*. C4ISRNET. https://www.c4isrnet.com/battlefield-tech/space/2024/03/27/space-force-to-upgrade-sensors-for-in-orbit-testing-training/

Amason, D. (2008). *SDO FlatSat facility*. Goddard Space Flight Center.

Bingen, K. A., Johnson, K., & Young, M. (2023). *Space threat assessment 2023.* Center for Strategic & International Studies.

Defense Intelligence Agency. (2022). *2022 challenges to security in space: Space reliance in an era of competition and expansion.*

Defense Science Board. (2021). *GEMS: Gaming, exercising, modeling, & simulation.* Department of Defense.

Department of Defense. (2007). *Directive 5000.59.*

Digital Twin Consortium. (2020, December 3). *Digital Twin Consortium defines digital twin*. https://www.digitaltwinconsortium.org/initiatives/the-definition-of-a-digital-twin/

Director, Operational Test & Evaluation. (2023a). *FY 2023 annual report.*

Director, Operational Test & Evaluation. (2023b). *Test and evaluation resources.*

Director, Operational Test & Evaluation. (2023c). *DOT&E strategy implementation plan—2023.*

Ender, T. (2014). Engineered resilient systems tradespace enabled decision making. *Engineered Resilient Systems Tradespace Enabled Decision Making*.

NASA. (2024, February 12). *State-of-the-art of small spacecraft technology*. https://www.nasa.gov/smallsat-institute/sst-soa/soa-communications/

National Institute of Standards and Technology. (2021). *Considerations for digital twin technology and emerging standards.* https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8356-draft.pdf

Office of the Undersecretary of Defense for Research and Engineering. (2023). *DoD instruction 5000.97 digital engineering.*

Rohde & Schwarz. (2023). *An overview of space electronic warfare.*

STORM, AFRL. (2023). https://intelligencecommunitynews.com/air-force-to-host-storm-industry-day-2023

Systems Engineering Research Center. (2019). *Technical report SERC-2019-TR-012.*

U.S. Army. (2024, April 2). *Test capabilities*. https://www.atec.army.mil/rtc/resources.html

Wright, M. (2008). Lunar reconnaisance orbiter FlatSat. *IEEE 13th European Test Symposium IEEE Computer Society and Test Technology Technical Council .* Verbania, Italy: Institute of Electrical and Electronics Engineers.

PANEL 20. ENHANCING ACQUISITION WITH ARTIFICIAL INTELLIGENCE

Thursday, May 9, 2024	
2:15 p.m. – 3:30 p.m.	**Chair: Rear Admiral Kurt J. Rothenhaus, USN,** Chief of Naval Research, Office of Naval Research ***Enhancing Acquisition Outcomes through Leveraging of Artificial Intelligence*** Ryan Novak and Rachel Giachinta, MITRE ***Improved Forecasting of Defense Acquisition Program Performance Using Digital Twin Models of a Revenue Centric Versus Cost Centered Approach (Enhanced Earned Value Management (E2VM))*** Raymond Jones, Naval Postgraduate School ***Introducing SysEngBench: A Novel Benchmark for Assessing Large Language Models in Systems Engineering*** Ryan Bell, Naval Postgraduate School

Rear Admiral Kurt J. 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.

From May 2020 until May 2023, he served as Program Executive Officer, Command Control Computers Communications and Intelligence (PEO C4I).

He assumed command of the Office of Naval Research as the 27th Chief of Naval Research in June 2023.

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.

He completed numerous acquisition tours including, program manager for the Navy's Tactical Networks Program Office (PMW-160) 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).

He is the recipient of the Legion of Merit, Meritorious Service Medal and various unit and sea service awards.

Enhancing Acquisition Outcomes through Leveraging of Artificial Intelligence

Christopher R. Barlow—is a seasoned Senior Acquisition Analyst with over 6 years of experience who has demonstrated expertise in audit risk mitigation, acquisition data analytics, process development and optimization, as well as demonstrated experience with independent assessments, challenge-based acquisitions (ChBA), artificial intelligence applications, and acquisition innovation. His diverse experience spans multiple federal agencies. Barlow holds a bachelor's degree in business administration from the University of Pikeville and is a Certified Business Intelligence Professional from Villanova University. [cbarlow@mitre.org]

Kevin M. Forbes—is a Principal Software Engineer at the MITRE Corporation, with over 20 years of software development experience. His expertise spans a wide range of technologies, including full stack software engineering, agile/DevOps, mobile apps, robotics, cloud computing and artificial intelligence. Forbes is deeply passionate about leveraging technology to address critical and urgent problems to have mission impact. His current work centers on the application of Generative AI to tackle challenges in government acquisition and software development life cycle. Forbes earned his bachelor's degree in computer science from Georgetown University and a master's degree in systems engineers from Johns Hopkins University. [forbes@mitre.org]

Rachel T. Giachinta—an early careerist passionate about applying cutting edge acquisition strategies to efforts that enhance U.S. access to, and competitive utility of, emerging technologies, Rachel Giachinta holds a bachelor's degree from the University of Virginia in global Middle Eastern and South Asian studies and a minor in global sustainability, laying the foundation for her interdisciplinary and critical thinking skills that she exercises daily at MITRE. With a diverse support portfolio across DoD and civilian agencies, Giachinta has supported both high-profile acquisitions and innovative pilot programs, approaching both as a mission driven problem solver. [rgiachinta@mitre.org]

Jay Kim—with an 8-year career in government and corporate contract acquisitions, Jay Kim has cultivated a rich array of experiences. His roles, including Contracting Officer Representative (COR), Contract Specialist, Contract Lead, and Acquisition Decision Analyst, have allowed him to refine his skills within the government sector and at the MITRE Corporation. Kim has made significant contributions to various aspects of government acquisitions, including acquisition planning and strategies, market and technical research, cost analysis, and contract negotiations and awards. He holds a master's degree in operations management, a bachelor's degree in political science, and an associate degree in accounting. [jaykim@mitre.org]

Zachary G. Levenson—is a Senior Contract Analyst at the MITRE Corporation who supports Intelligence Community, Defense, and Federal sponsors throughout their acquisition life cycles. His interests include security and counterintelligence risk associated with the acquisition and contracting practices of the federal government. He has 5 years of experience in acquisition and contracting and is a former Subcontract Administrator. He holds an BA in political science from West Virginia University and a Certificate in Contract Management from Villanova University. [zlevenson@mitre.org]

Ryan M. Novak—acquisition Outcome Lead from the MITRE Corporation and AI Lead for Acquisition Innovation Center, has earned two MBAs, an MS in strategic purchasing, and an MS in project management, and is DAWIA III Contracting certified. He has authored the Challenge-Based Acquisition Handbook, taught the approach to 500+ staff at MITRE, Industry, and the Government, and helped numerous agencies employ innovative acquisition approaches. He is a former USAF Contracting Officer with 28 years of leadership experience in the areas of innovation, contracting, acquisition, negotiation, strategic purchasing, consulting, and project and program management. He is a published author, speaker, and practitioner on innovative acquisition approaches and solutions. He helped achieve strategic program and acquisition successes for the DoD, Civilian Agencies, and the IC across the board. Novak led the start-up on iCAMS, has held an unlimited CO warrant, and completed both the Executive Strategy Program at Oxford and the Accelerated Management Program from Yale University. [rnovak@mitre.org]

Justin G. Raines—is a Principal Decision Analyst at the MITRE Corporation, where he assists government teams to develop effective acquisition strategies. His work focuses on developing innovative capability and industry communication in order to improve product delivery and meet national priorities. Raines has extensive experience in nontraditional procurement strategies, challenge-based acquisition, market research, and strategic planning. He is currently an Acquisition Program Manager in the USAF Reserves within the Kessel Run program office; he holds a bachelor's from the USAF Academy and a Master of Public Policy from the University of Maryland. [jraines@mitre.org]

Stephen W. Roe—managing Director at MITRE for Veterans Affairs programs since 2020, has significantly contributed to advancing the VA's strategic initiatives by generating cross-federal collaboration opportunities and developing several agile teams across VA's field offices. With a MITRE tenure starting in 2004, his efforts have enhanced system strategies and business transformations within the VA and beyond. Previously, as Portfolio Manager and National Background Investigation Services (NBIS) Project Leader, he led a 45-member team on a $15 million budget DevSecOps project. Roe also co-authored MITRE's Challenge-Based Acquisitions (ChBA) Handbook. He holds two systems engineering degrees from Virginia Tech, Johns Hopkins University, and an MBA from the University of Virginia's Darden School of Business, and is the Pride Council Business Resource Group Co-Chair. [sroe@mitre.org]

Abstract

The extraordinary advancement of Artificial intelligence (AI) technology emerges at a critical juncture in which the Federal acquisition workforce is ill-equipped to meet the sky rocketing demand for products and services, alike. AI poses the opportunity to overcome data-intensive, laborious tasks and expedite the speed in which acquisition professionals operate; potential benefits may increase efficiency, enhance transparency, and reduce workload. While the use of AI across the Federal Government differs between agencies, the significance and scrutiny of Government Acquisition makes implementing AI across the acquisition process uniquely challenging. This paper will explore the current state of AI; who (i.e., which agencies) and how AI currently supports the acquisition process across the Federal Government. Next, the future state of AI and anticipated applications for the acquisition community will be discussed…think the future, think the next generation of Acquisition! This will be developed through strategic exploration across thought leaders, academic research, and working within our own AI model for acquisition. Next, we will discuss how the risks of this new technology -- new tools and novel concepts -- introduce both procedural, ethical, and operational risks that must be taken into consideration. Finally, we will offer a set of recommendations on how best to implement AI in the acquisition process as well as a list of best practices to maximize utility, mitigate risks, and ensure the acquisition workforce is well positioned to embrace the benefits and efficiencies of integrating AI capabilities.

Introduction

Artificial intelligence (AI) has stimulated widely divergent popular opinions and is a topic at the forefront of thought leaders' attention. As the world grapples with understanding AI's potential impacts, especially in augmenting or eliminating existing processes, these technologies continue to gain prominence. So-called "generative AI" has captured attention due to its ability to replicate human talents such as producing text, audio, and imagery. AI takes only seconds to perform activities that can take humans years to master. Beyond its "Wow" factor, generative AI has the potential to streamline labor-intensive and tedious tasks.

This paper explores the opportunity that judicious use of AI provides in reducing or eliminating tedious, labor-intensive acquisition tasks that often divert attention from strategic planning and execution. While the acquisition profession has already embraced technology, the Federal Acquisition process remains notoriously slow in acquiring and ultimately delivering goods, services, and/or capabilities to those who need them.

This paper does not propose a future in which AI completely replaces human involvement in acquisition activities. Instead, it investigates how AI can serve as a supportive tool that augments the workload of basic, routine, standardized actions and, thus, enables Federal Acquisition to operate more efficiently and rapidly. It envisions a balanced integration of AI in acquisition by leveraging AI's potential to enhance human productivity and decision-making, thereby increasing efficiency without disregarding human oversight and expertise.

Acquisition often views itself as more of an art than a science, valuing the unique perspectives that individuals contribute to the decision-making process. Within this framework, it is understood that certain parts of the acquisition process should continue to benefit from individuals' expertise. However, some portions of the process involve extensive manual handling of data, which can distract acquisition experts from concentrating on overarching strategic objectives. This paper envisions a world in which the art of acquisition is blended with scientific components of AI, enabling acquisition professionals to focus, prioritize, and work smarter, not harder.

Unraveling how the Federal Acquisition process uses AI today reveals that some acquisition professionals acknowledge its value and have adopted and implemented the technology, despite the perception of AI's relative immaturity. To accomplish this, the paper begins by examining the current state of AI in the Federal Acquisition community, including insight from early adoptions and lessons learned. The discussion then shifts to the future, as envisioned by technologists and forward-leaning acquisition professionals. The discussion identifies tangible points where AI can enhance and expedite the acquisition lifecycle.

Finally, the paper contains recommendations on what acquisition professionals should consider or do today. Whether they fear or overestimate and potentially even sensationalize the impact of AI, acquisition professionals must rid themselves of common misconceptions around AI, such as the belief that AI will have an immediate and overwhelming impact on their work environment. Like it or not, the transition to AI has already begun. Organizations leading this shift are proceeding cautiously with development, testing, and integration. Many of the personnel within these flagship organizations may still be unaware of AI's inevitable integration into their workflows. This indicates that new adopters must integrate AI into their operations gradually, building employee trust and confidence in AI security and the validity of its outputs, while maintaining the "trust but verify" model. This process will require strategic planning and careful communication. As Executive Order (EO) 14110 states, "Artificial Intelligence must be safe and secure. Meeting this goal requires robust, reliable, repeatable, and standardized evaluations of AI systems, as well as policies, institutions, and as appropriate, other mechanisms to test, understand, and mitigate risks from these systems before they are put to use" (The White House, 2023c).

Understand the Nomenclature

Before exploring the impact of generative AI on the Federal Acquisition lifecycle, one must first understand what AI *is* and *is not*. The extraordinary speed at which AI has emerged and its wide-ranging applications have led to a proliferation of definitions, potentially hindering practical, consistent, and regulated implementation. Acquisition professionals should consider this section as a reference source on the essential attributes of AI!

According to the American National Standards Institute (ANSI) International Committee for Information Technology Standards (INCITS), AI is "a branch of computer science dedicated to creating systems that perform task associated with human intelligence, like reasoning, learning, and self-improvement." Generative AI—the focus of this paper—is a specific branch of AI, akin to how engineering encompasses various specialties such as mechanical, electrical, etc. It is a type of artificial intelligence that can learn from and mimic large amounts of data to create content such as text, images, music, videos, code, and more, based on inputs or prompts (Harvard University Information Technology, n.d.).

The overall discipline of AI has subfields:

- **Machine learning (ML)** is the capability of a machine to imitate intelligent human behavior, enabling computers to learn without explicitly being programmed (MIT Sloan, 2021).
- **Natural language processing (NLP)** combines computational linguistics—rule-based modeling of human language—with statistical, machine learning, and deep learning models to give computers the ability to understand text and spoken words in much the same way human beings can (IBM, n.d.).
- **Large language models (LLMs)** "are neural network models designed to process sequential data (e.g., can be trained by giving it access to a large corpus of text (such as Wikipedia, digitized books, or portions of the Internet) and using that input text to learn to predict the next word in a sequence, given what has come before" (Brynjolfsson et al., 2023).

Before making or implementing policy, decision-makers must first understand this basic taxonomy; it underpins a comprehensive understanding of how AI works, and it informs the ways that AI can be effectively and safely harnessed.

Current State: AI in Federal Acquisition in the Year 2024

The legislative and executive branches of Government have recently focused significant attention on AI policies. The 117th Congress, for example, introduced 75 AI-focused bills, of which six (6) were enacted (Congressional Research Service, 2023). From initiatives such as the National Artificial Intelligence Initiative (NAII) Act of 2020, which aims to codify and support American AI investment and advancement, to EO 14110, which has the primary objective of regulating that advancement amidst rapid expansion, official guidance has attempted to unify the fragmented approach to developing and deploying AI technology.

By design, legislation and executive branch policy documents often provide conceptual rather than practical guidance on how government organizations should leverage and develop AI for their specific missions. Examples include the National AI R&D Strategic Plan (The White House, 2023b), National AI Research Resource Roadmap (The White House, 2023a), and the Office of Management and Budget's Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence (Young, 2023), although other AI-based policies abound. This absence of detailed direction creates barriers to action. Identifying these gaps can aid acquisition organizations in developing future guidance and approaches for integrating AI that enhances the Federal Acquisition process.

Across the Federal Government, various working groups, governance bodies, and agency initiatives provide disjointed, local guiding principles, scope, frameworks, and tools. In order to implement the AI-driven acquisition enhancements of tomorrow, today's acquisition professional must transcend these silos and understand insights and contributions from today's AI pioneers.

Fundamentally, AI is no longer a theoretical technology, poised to shape our future defense initiatives; rather, it is already here, influencing the acquisition process in a host of ways. Acquisition leaders must accept this and securely embrace the power that AI brings. While the future of what AI can do remains relatively opaque, its immediate impact on the acquisition process can already be imagined. Interviews by this authorship team with acquisition thought leaders and AI technologists revealed potential "quick win" areas for AI in Federal Acquisition. To demonstrate the power of AI, the authors of this paper cross-referenced the collected feedback with a ChatGPT-generated list of "how the Federal Government should use AI/ML in Procurement" (DelTek, 2023). Table 1 presents the ChatGPT-suggested areas, with green boxes highlighting those also identified by interviewed subject matter experts (SMEs).

Table 1. AI-Identified Areas for AI Application in Federal Acquisition

Identified only by SMEs	Identified by Both SMEs and ChatGPT	Identified only by Chat GPT
Requirements Development	Supplier Performance Monitoring	Market Intelligence
Drafting Contract Language	Vendor Selection	Predictive Maintenance
Commercial Analysis Software for Drafting Clauses in Compliance with Plain Writing Principles	Contract Compliance Monitoring	Demand Forecasting
Drafting Statutes and Regulations	Spend Analysis	Risk Assessment
Developing Evaluation Criteria	Contract Management	Fraud Detection
	Workflow Automation	

Table 1 does not suggest that AI should be the arbiter of how it is applied, as shown by its omission from the possible application areas identified by SMEs. Nevertheless, the green areas in Table 1 indicate that AI can aid in brainstorming ideas that human SMEs can refine and expand upon.

Industry has also recognized contract management and consistency as key areas in which AI can augment human skills. The *Harvard Business Review* (Rich, 2020) reports that AI can reduce human error in contract management and compliance monitoring by identifying and extracting key data points. Feeding historical contracts into AI systems could help establish performance standards or evaluation criteria. However, human oversight is essential to verify the relevance of AI-generated wording and to capture a program's contractual nuances.

Some of these techniques are indeed being applied to the Federal Acquisition process today. Table 2 shows a sampling of the current use of AI tools in Federal Acquisition, serving as proofs of concept for the potential applications detailed above. This should inspire, not limit, ideas for broader AI adoption. While these tools provide foundational capabilities, organizations should customize tools to meet specific Federal Acquisition needs, drawing on lessons from these early use cases to improve future iterations.

Table 2. Current State of "Who's Using What" AI in Federal Acquisition

Organization	*Tool*
Department of Agriculture (USDA)	*Acquisition Approval Request Compliance Tool*. Utilizes the text in the procurement header and line descriptions within USDA's Integrated Acquisition System (IAS) to determine the likelihood that an award is IT-related and therefore might require an After-Action Report (AAR). The model uses the text characteristics for awards that have an AAR number entered into IAS and then calculates the probability that those procurements that did not have an AAR Number entered in IAS are in fact IT related (USDA, n.d.).

Organization	*Tool*
Health and Human Services	*Federal IT Acquisition Reform Act (FITARA) Tool*. Automates the identification of National Institute of Allergy and Infectious Disease (NIAID) IT-related contracts (HHS, 2023).
Treasury Department	*Digital Accountability and Transparency Act (DATA) Bot*. Automates the verification process of reports coming from the Internal Revenue Service (IRS) Federal Procurement Data System (FPDS) with the information in contract documents. Leverages NLP to extract unstructured information from contract documents and uses F1 scores to measure performance of validation models for each specific data element (U.S. Department of Treasury, 2024).
Department of Labor (DoL)	*Intranet Website Chatbot Assistant*. Uses a conversational chatbot on DoL intranet websites to help answer common procurement questions as well as questions about specific contracts (DoL, n.d.).
Department of State – With 37 reported AI use cases, two serve the acquisition process.	*Federal Procurement Data System (FPDS) Auto-Populate Bot*. Automates the data entry in FPDS to reduce the burden on the procurement staff and drive improved compliance on DATA reporting. It is used to update ~300 FPDS awards per week.
	Production Service Code Automation ML Model. Scans unstructured, user-entered procurement data (e.g., requisition title, line descriptions) to automatically detect the commodity and service types being purchased and thus enhance procurement categorization (Department of State, 2023).
General Services Administration (GSA)– With 12 reported AI use cases, five pertain specifically to the acquisition process (*AI Inventory — Tech at GSA*, n.d.).	*Solicitation Review Tool (SRT).* Pulls in SAM.gov data about solicitations, then compiles the data into a database for use by ML algorithms. An NLP model determines if a solicitation contains compliance language. If a solicitation does not include compliance language, it is marked as non-compliant. Each agency is asked to review its data and validate the SRT predictions. The GSA also conducts random manual reviews monthly.
	Acquisition Analytics. Takes detailed data on transactions and classifies each transaction within the Government-wide category management taxonomy.
	Category Taxonomy Refinement. Uses NLP to extract tokens from product descriptions more accurately to shape intended markets for Product Service Codes (PSCs).
	Contract Acquisition Lifecycle Intelligence (CALI). Streamlines the evaluation of vendor proposals against the solicitation requirements to support the Source Selection process using an automated ML evaluation tool. CALI is currently being trained with sample data from the End User License Agreements (EULAs) under the Multiple Award Schedule (MAS) program.
	Chatbot for Federal Acquisition Community. Enables the GSA FAS NCSC (National Customer Support Center) to streamline the customer experience process and automate documentation of answers to commonly asked questions through public-facing knowledge articles. The end goal is to reduce staffing requirements for NCSC's live chat programs and allow NCSC resources to be dedicated to other proactive customer service initiatives. Customers still have the option to connect to a live agent if they choose by requesting an agent.

Organization	*Tool*
DoD's Chief Digital and AI Office (CDAO)	*Acqbot*. Like ChatGPT, generates text to accelerate authorship of acquisition artifacts (e.g., problem statement; Heckman & Heckman, 2023).
Department of the Army	*Determination of Responsibility Assistant (DORA)*. Pulls information from SAM.gov and the Federal Awardee Performance and Integrity Information Systems (FAPIIS) to deliver relevant information to Contracting Officers (COs; i.e., as specified by inputting a vendor's DUNS number), rather than having COs access each system separately. Within minutes, the DORA bot sends COs a summary document of the vendor's responsibility status and results, including screenshots of what the vendor's file contains (Kanowitz, 2023).

Executive Order 13960, "Promoting the Use of Trustworthy AI in the Federal Government" (Federal Register, 2020) plays a key role in making AI usage data accessible by requiring Federal agencies to document AI use cases; a webpage has been stood up to communicate these to the public to ensure transparency (The White House, n.d.). However, SMEs suggest that these documented use cases may not capture all AI explorations being pursued by Federal Acquisition professionals. For example, the MITRE Corporation is working on an LLM that could function as an interactive Federal Acquisition Regulation (FAR) bot and one that can quickly assess market research. Wolf Stake has also developed prototypes to assist contracting efforts. Government may have been pursuing similar efforts, but no documented cases exist (Wolf Stake, n.d.). Despite the availability of the Federal inventories, the Federal Acquisition community has not catalogued, or shared lessons learned to aid broader AI adoption, integration, and/or training.

A growing trend in government is the development and use of AI chatbots that function like help desks but also assume a small degree of AI discretion (i.e., complete full tasks). Agencies have used these chatbots to perform simple actions such as data summarization and key word searches through records. The DoD's Defense Information Systems Agency (DISA) has considered a 'digital concierge' LLM to assist employees "in all aspects of their job." DISA estimated that about 80% of the data analyses currently performed by defensive cyber analysts could be automated, freeing up human resources for complex tasks so that "their brains can be applied to those really high-end problems" (Gill, 2023). This concept could also benefit acquisition professionals, automating tasks such as market research.

Future State: 20 Acquisition AI Use Cases for the Year 2035

The Acquisition Management Landscape

Government acquisition managers are intimately familiar with the so-called 'Iron Triangle' of cost, schedule, and performance. Successful acquisition management requires balancing the three. Moreover, acquisition managers are deeply aware that perturbations to one element of that triangle have impacts on the other two. For acquisition managers, applying AI fundamentally distorts the dynamics of the Iron Triangle. It becomes possible to perform their tasks faster or less expensively without compromising the other elements. While AI can never fully replace the talent of an acquisition management team, it can be used across the acquisition lifecycle to augment their efforts and improve outcomes.

20 AI Use Cases for Acquisition

Consider a thought experiment. The year is 2035. Use of AI has become ubiquitous. It permeates the daily operations of the acquisition management team. Imagine, then, the work of the Integrated Product Team going through the end-to-end acquisition and contracting lifecycle to procure a new system or service. How does the team use AI?

The following depicts the outcome of such a thought experiment. While a subsequent section of this paper describes those threats and drawbacks in detail, it is important to note for the purposes of this thought experiment that acquisition programs can manage those risks successfully. The potential applications are nearly limitless; however, those identified here may be likely candidates for implementation in the near-term.

1

Building a Request for Information (RFI)

Detailed Description: As the acquisition team begins to fully understand and address the agency's need, it recognizes the need to engage industry immediately. Knowing that more detailed, face-to-face market research will occur later, the team wants to gauge initial industry interest and "big picture" suggestions. Applying AI, the team rapidly creates an RFI that draws upon historical agency documentation, the acquisition team's existing products (e.g., briefings to leadership), and notes from several brainstorming sessions. A second AI-enhanced tool reviews the AI-generated RFI to ensure the RFI does not contain controlled unclassified and proprietary information, compartmentalizing different tasks across different AI tools. The acquisition office uses different AI systems to perform different tasks, demonstrating the limitations of a single AI system to accurately execute all tasks. The acquisition team then reviews and enhances the RFI, enabling rapid release to industry.

Impact on Acquisition Outcomes: A process that used to take 6 to 8 weeks is shortened to 10 days, accelerating schedule and enabling the program to perform other, more detailed vendor engagements without adversely impacting schedule.

Primary Improvements:	Cost	Schedule	Performance

2

Industry Constructing Responses to Government

Detailed Description: Each prospective vendor's AI system examines the Government's front-end portal on a daily basis, generating a 'score' for each newly posted solicitation or request; past performance criteria may influence the scores. This AI tool informs vendors' capture managers about opportunities, reduces the amount of labor required to find opportunities, and allows industry to better invest scarce business development resources in the most potentially productive lines of effort. After identifying a highly scored RFI (i.e., one that might lead to contract award), vendors use AI applications to respond to the Government's request. By leveraging existing capability statements, project information, resumes, and other corporate data, the AI system can rapidly produce a high-quality response that includes several key innovations the company has produced for the Government to consider.

Impact on Acquisition Outcomes: Industry only needs to dedicate small amounts of time, resources, and effort to produce high-impact and meaningful RFI responses, including ones that give game-changing technical and strategy recommendations. This has two positive impacts. First, vendors can reduce their overhead rates, thus increasing competitive pricing for future opportunities. Second, vendors become more willing to engage in one-on-one and in-depth market research activities with the Government, as responding to the RFI did not consume their scarce resources.

Primary Improvements:	Cost	Schedule	Performance

3

Building Market Research Assessments

Detailed Description: The Government acquisition office uses AI tools to examine all industry responses to the RFI. The AI system can search through all the elements of each response, determine overall capabilities of different companies, and provide insight into the quality of the response, as envisioned in the RFI. The AI system's reports allow the acquisition office to more rapidly assess RFI responses, attach meaning to the different recommendations from industry, and efficiently gain acquisition insights. Specifically, the AI assessments enable the Government to prioritize its actions prior to formal solicitation and weigh different industry recommendations.

Impact on Acquisition Outcomes: Often, Government acquisition offices do not make full use of industry input (e.g., RFI responses) during the market research phase – they do not read the responses in detail and sometimes overlook specific recommendations. The AI-assisted reports on each response focus on critical information, potentially avoiding future risks and accelerating acquisition delivery timelines.

Primary Improvements:	Cost	Schedule	Performance

4

Performing Open-Source Market Intelligence

Detailed Description: The acquisition office uses an AI tool to ask pointed, capability-driven questions. Specifically, the program seeks to understand what potential industry partners that are not part of the typical Government contracting ecosystem may be qualified to bid for a particular contract. The AI produces a summary report for the acquisition office, which uses it to quickly understand the state of practice in industry, including pricing models. The acquisition office then publishes an open RFI on SAM.gov. Once approved by the acquisition office, the AI sends an email to each qualified company's cognizant business office recommending that the company consider responding to the open RFI.

Impact on Acquisition Outcomes: Acquisition offices always welcome better information about leading-edge capabilities available from industry. Increasing participation by all segments of industry in the pre-solicitation process expands competition to ultimately drive down cost, improve schedule, and enhance vendor performance.

Primary Improvements:	Cost	Schedule	Performance

5

Assessing Risk of Existing Efforts

Detailed Description: Ongoing contracts demonstrate varying degrees of successes and failures across a multitude of cost, schedule, and technical performance metrics. Monthly program management reports, irregular reports and briefings to agency leadership, corrective action reports to contractors, cost and financial status reports delivered under contracts, user feedback on deployed capabilities, and, in extreme cases, legal records contain evidence of such outcomes. The acquisition office tasks a custom AI tool to examine all of these data sources in detail and identify (1) potential areas of risk or failed performance that may be relevant to the current acquisition/contract and (2) potential metrics, program management approaches, and insights that enabled successes in similarly scoped programs. The acquisition office uses the AI-generated report to add and modify contract requirements and internal program management processes.

Impact on Acquisition Outcomes: Learning from the successes and failures of other programs is critical for a successful acquisition. Acquisition offices often write 'lessons learned' that frequently become no more than 'lessons documented' – acquisition offices simply do not have the time or resources to fully understand and internalize the experiences of other programs. AI can reduce the need for such resources and the effort involved in reviewing the historical record, allowing current acquisitions to connect to valuable lessons from the past. This increases the likelihood that the acquisition will not repeat past mistakes and will repeat successful approaches.

Primary Improvements:	Cost	Schedule	Performance

6 **Identifying Similar Programs**

Detailed Description: The program office tasks its AI tool to comb SAM.gov, Government spending data, and contract repositories from across the Federal Government. The tool makes data readily available and informs the acquisition office of other agencies and departments that are acquiring similar (or the same) capabilities. Communication across organizational boundaries and stovepipes can be extremely challenging for Government agencies. The nature of Government operations makes it difficult for agencies to gather current or complete information on the activities of all other departments and agencies. The AI tool can bridge the communication gap when given access to data generated by various agencies and provide detailed insight into similar acquisitions by other organizations.

Impact on Acquisition Outcomes: Using the AI-generated information, acquisition offices can access information about acquisitions conducted by other agencies, learn from these various sources, and potentially leverage their existing contract vehicles. They can use this information primarily to enhance affordability and improve technical performance by leveraging existing capabilities.

Primary Improvements:	Cost	Schedule	Performance

7 **Developing Analyses of Alternatives**

Detailed Description: Using market research, open source, and internal agency data, the AI system can perform detailed course-of-action analyses for the proposed acquisition. Specifically, it develops informed acquisition strategy alternatives, including contracting options, and delivers technical alternatives for the acquisition team to consider. By pairing these different options with associated risks and opportunities, the AI system gives the acquisition office in-depth insight into different approaches that it can leverage. Moreover, the AI system may identify alternatives that acquisition staff may otherwise have overlooked.

Impact on Acquisition Outcomes: Acquisition offices should make decisions based on the greatest possible amount of information. They can use AI to organize available information, better structuring an acquisition manager's decision space. While AI should not make independent recommendations, it can inform available options from the risk and opportunity perspectives, positively impacting all elements of program success.

Primary Improvements:	Cost	Schedule	Performance

8 **Building Management Plans**

Detailed Description: Procedurally, acquisition offices expend many resources on building, refining, and reviewing documentation. While the direct utility of many documents may vary by program, the need for such documentation is rarely in question. The law directs programs to develop, and follow, plans that describe their processes and procedures. Often, documentation takes the form of a Systems Engineering Plan, a Project Management Plan, and/or a Test and Evaluation Plan. The program office uses AI to produce these required plans. Starting with past approved plans, augmenting them with the team's copious notes as to how the acquisition office will manage the program, and ingesting recordings of team meetings, the AI system generates partially complete plans. It fills in the areas that it can from the source material and highlights missing material. The acquisition team can use this material as a starting point for developing the required documents, reducing the time and complexity of such undertakings by freeing the team to spend more time on developing new concepts and less time on writing.

Impact on Acquisition Outcomes: High-skilled acquisition labor is best suited to innovating, analyzing, and synthesizing information. Writing routine program documentation can often distract from the acquisition team's most valuable work. Using AI to help produce some of this documentation frees the acquisition team to focus on more important tasks.

Primary Improvements:	Cost	Schedule	Performance

9 **<u>Creating Work Statements and CDRLs</u>**

Detailed Description: Many components of contractor requirements, such as the work statement (e.g., Performance Work Statement or Statement of Work) and Contract Data Requirements Lists (CDRLs) include basic, repeated, or high-level requirements that can be derived from existing sources. Additionally, those documents often incorporate requirements, or parts of requirements, from prior contracts. The AI system produces first drafts of these critical contract documents for the program office to consider. In the course of doing this, the AI can determine 'best of breed' requirements from different sources of materials and create a superior document. This allows the program office to focus its time on innovations, key requirements, and other differences from status quo.

Impact on Acquisition Outcomes: This AI implementation primarily affects the schedule-to-release time for generating solicitation documentation, enabling the Government to have a better set of documents earlier. Additionally, the quality of the documents is likely to improve, as they will incorporate advances and recommendations across the breadth of requirements.

Primary Improvements:	Cost	Schedule	Performance

10 **<u>Estimating Cost</u>**

Detailed Description: Using an extensive database of market research, open source data, agency contract data, work statements, and financial data, the AI system builds a comprehensive cost estimate for the acquisition. Each calculated element includes a tailored confidence level and highlights areas with insufficient data to accurately calculate an estimate. The program office uses this AI-generated estimate to start its costing and make well-informed trade-off decisions.

Impact on Acquisition Outcomes: The AI-informed cost estimate more accurately reflects reality and enables the program office to procure the right solution at the right price.

Primary Improvements:	Cost	Schedule	Performance

11 **<u>Building Contract Considerations and Recommendations</u>**

Detailed Description: Using an extensive library of case law, protest responses, and previous contracts, the AI tool analyzes the Government's draft solicitation and model contracts. It seeks out potential vulnerabilities, protest grounds, and areas that require clarification(s). Furthermore, the AI tool makes recommendations on how to improve the solicitation and model contract. The CO then uses this report to bolster the solicitation and improve the overall quality of the Government's contract.

Impact on Acquisition Outcomes: While the improved contract does not directly affect the schedule, cost, or performance of the acquisition, implementing this AI use case does reduce the likelihood of a sustainable protest and increases the clarity of the solicitation. Ultimately, this may reduce the need for re-work and protest-imposed pauses as well as improving the quality of responses.

Primary Improvements:	Cost	Schedule	Performance

12 **<u>Tailoring Training for New Team Members</u>**

Detailed Description: The acquisition team brings new technical SMEs and functional teammates on board throughout the pre-solicitation period. To rapidly bring these new team members up to speed, the AI system produces a custom, high-quality training video for each team member, including the information most relevant to each person's area of expertise. The system asks each team member to fill out a survey regarding his or her experience with the acquisition, agency, technology, and other factors. It compiles programmatic documentation, slides, presentation videos, and other agency information into a succinct 6-hour training video. This video addresses key elements of the acquisition in question. For members who are new to the agency, the video includes detailed background on the agency and its mission. For those with expertise in a specific domain (e.g., software engineering), the video describes relevant capabilities sought and summaries from market research.

Impact on Acquisition Outcomes: The AI-generated training produces high-quality materials that enable new team members to attain full effectiveness. It does so without burdening existing staff members and allows new staff members to rapidly begin contributing to programmatic outcomes.

Primary Improvements:	Cost	Schedule	Performance

13

Compiling a Comprehensive Bidders Library

Detailed Description: Offerors require detailed information on existing capabilities, contracts, needs and requirements, priorities and challenges to provide the best quality and risk-informed proposals. Information is key to reducing uncertainty and de-risking contract performance. In general, the more information a potential vendor can apply to its proposal, the better that proposal and the subsequent capability will be. The acquisition office tasks its AI system to curate a library of information for the contractor community. The system includes all relevant information, filters out all sensitive information, flags any conflicting information for program office action, and neatly indexes the content for offeror consumption. The AI system organizes the bidders' library in a logical flow, allowing vendors to more rapidly understand the information that the library contains.

Impact on Acquisition Outcomes: Creating a comprehensive and impactful bidders library can be challenging for program offices. Moreover, organizing that information into a logical index consumes a considerable amount of time. By using AI, the acquisition office can improve the quality of proposals and the product, service, or /capability to be delivered, positively affecting cost, schedule, and performance during execution.

Primary Improvements:	Cost	Schedule	Performance

14

Checking Proposals for Compliance

Detailed Description: The AI system compares each received proposal to the Section L or the instructions to offerors that accompanied the solicitation. In this analysis, the AI system scores each proposal to determine if any proposal elements are missing, if the elements contain insufficient information, or if pieces of the proposal fail to meet Government requirements. The AI system generates a report that the contracting office can then use to verify if the proposal(s) in question fails to meet compliance requirements.

Impact on Acquisition Outcomes: This use of AI helps the source selection team perform its tasks faster, reducing the need to analyze deficient proposals and reducing the number of clarifications that the agency must provide to industry.

Primary Improvements:	Cost	Schedule	Performance

15

Industry Generating Proposal Content

Detailed Description: After receiving the solicitation, offerors use a variety of AI tools to build the basis of their proposals. Specifically, the AI system creates the basis and baseline of the proposal by comparing the solicitation requirements to other, similar proposals that the company has submitted previously. It also relies on a library of information about each vendor, including capability statements, RFI responses, and other business development and technical implementation material. While the data does not provide enough information for the AI system to formulate a comprehensive proposal, it does provide a starting point. Once the proposal managers and technical team have finalized the proposal, they use a separate AI tool to compare the proposal to the Offeror Instructions and Evaluation Criteria in the Request for Proposals (RFPs). The tool provides recommendations on how to improve the clarity and effectiveness of the proposal.

Impact on Acquisition Outcomes: Commercial companies' use of AI to build proposals enhances the quality and readability of vendor proposals. Furthermore, it allows vendors to focus their limited business development resources on presenting the best approach to solving the most complex technical problems. Ultimately, this ensures that the Government obtains the best innovations and technical solution within the limited time available for bidders to submit their proposals.

Primary Improvements:	Cost	Schedule	Performance

16

<u>Conducting Initial Proposal Scoring</u>

Detailed Description: Given evaluation criteria and a set of proposals, the acquisition office can use AI to perform initial proposal analysis. Specifically, an AI system can compare proposals to the solicitation requirements and evaluation criteria. The resulting report can identify risk hot spots to which the evaluation team should pay especially close attention and identify areas of each proposal that may have failed to meet requirements. The acquisition office can also use the AI system to inform relevant exchanges with industry, identifying inconsistencies within the proposal or illogical statements.

Impact on Acquisition Outcomes: Source selection is a time-intensive process, often requiring months to explore and evaluate comprehensive offeror-provided information, compare it to evaluation criteria, and determine the Government's best option. Using AI to perform some initial proposal scoring can accelerate this process and make it less laborious. While AI will never replace the role of a source selection evaluation team or technical evaluation team, an AI-based system can help streamline the process and augment such efforts.

Primary Improvements:	Cost	Schedule	Performance

17

<u>Building Source Selection Technical Reports</u>

Detailed Description: The source selection's technical evaluation team has reviewed every proposal. The team has created individual notes on the proposals using templates that focus on risks and opportunities. For consensus events, the team members have created consolidated notes and recorded their conversations. The AI system uses all these sources of information to compile a comprehensive first draft of the technical evaluation report. It compares the source data to the evaluation criteria in the RFP and the content of the proposal. It accurately cites all solicitation and proposal references. It highlights areas of the evaluation that do not make sense or content that is potentially inconsistent with the solicitation or irrelevant. The evaluation team reviews this draft document, adjusts it appropriately, and finalizes it to state the agency's official position. The AI system does not create new evaluation material; it simply condenses the existing evaluation data into a readable technical evaluation report.

Impact on Acquisition Outcomes: Source selection teams spend a tremendous amount of time organizing thoughts, finding references, and documenting their findings. Use of AI accelerates this process and allows the teams to spend more time focusing on the proposal content rather than on how to document it.

Primary Improvements:	Cost	Schedule	Performance

18

<u>Compiling Past Performance Data</u>

Detailed Description: The source selection team needs past performance data to evaluate each vendor. For each offeror that submits a proposal, the AI system uses the Contractor

Performance Assessment Reporting System (CPARS) and other contract performance data to form comprehensive past performance reports on each company. Furthermore, the AI system automatically reaches out via email to program managers across Government who might potentially have input on the past performance rating. Those program managers then submit their input to the AI database, which uses the inputs to enhance the past performance findings. The AI system compiles a comprehensive past performance report for the source selection evaluation team to consider as it forms its ratings.

Impact on Acquisition Outcomes: AI can consolidate much of the work needed to support an evaluation of past performance. Specifically, the AI system compiles a past performance report without human intervention, enabling the Government to make better use of its scarce acquisition resources and focus only on generating ratings from the gathered information.

Primary Improvements:	Cost	Schedule	Performance

19

Accelerating Transition-In

Detailed Description: Once the contract has been awarded, the awardee uses an AI-based tool as a primary resource to begin work under the contract. Specifically, the contractor relies on the AI tool to help identify the right staff to bring onto the team – both current employees of the company and possible new hires. The AI system uses contract and corporate information to draft initial reports, such as program management plans, risk management plans, monthly reports, and other necessary documents. The system drafts initial kick-off slides for the contractor's program team to consider. In each case, the AI system seeks to reduce the labor and risk required to complete the project.

Impact on Acquisition Outcomes: Using AI in this way allows contractors to perform their work more efficiently and at a reduced cost. It frees the contractor's team to focus less on administrative work and more on higher-risk elements of contract performance. This reduces the burden imposed by routine programmatic work and allows the contractor to address higher-risk technical aspects of the work more thoroughly.

Primary Improvements:	Cost	Schedule	Performance

20

Detecting Performance Assessment Hot-Spots

Detailed Description: The Government program office's AI system uses programmatic data, including contractor program reviews, contract data requirements list (CDRLs), and meeting notes to create a comprehensive program risk register. This register provides the Government acquisition team with a detailed view of the program's cost, schedule, and technical performance variables. Moreover, the AI system helps prioritize the higher risk items and draws attention to those areas that require Government action. While the AI system does not replace the need for diligent programmatic oversight, it enhances the Government team's ability to rapidly and thoroughly process the vast quantities of information generated by the program and its contractors.

Impact on Acquisition Outcomes: It is difficult for acquisition teams to rapidly process the vast quantities of information generated by any given program. Judicious use of AI helps acquisition managers to gain timely insight into the risks and opportunities that permeate their program. This helps the Government team to manage performance while simultaneously enabling the team to focus on the highest priority items.

Primary Improvements:	Cost	Schedule	Performance

While it is important to note that each of these use cases is notional, the examples all demonstrate "the art of the possible" that AI may soon deliver to the acquisition workforce. AI can complement the detailed work that acquisition teams must carry out, augmenting human intelligence, labor, and process flows, and synergistically improving the overall quality and timeliness of the acquisition process. The structure to implement this AI-driven future is already here. Acquisition leaders must learn to embrace how AI will change their functions, the functions of their teams, and the ways in which they build and buy tomorrow's key technologies.

Before these ideas can become reality within the field, acquisition offices must be open to this new world, be avid learners and early adopters, and ultimately, and most importantly, address risks and challenges. While the opportunities for AI-based improvements are nearly limitless, practitioners must recognize, and mitigate, the associated threats and risks.

AI Threat Landscape – A Manageable Risk Today and Tomorrow

It is clear that AI can transform the processes used by the acquisition community. AI-enhanced applications offer many opportunities to make the acquisition workforce more efficient, decrease individual workloads, and decrease Procurement Action/Administrative Lead Time (PALT). However, while it can bring huge benefits, AI also poses additional risks.

By analogy, the internet offers tremendous advantages to the acquisition process; however, like AI, threat vectors abound across the internet. Universal reliance on the internet has made contractors and the Government increasingly susceptible to attack from any location and by any adversary around the world. Instead of eschewing the internet, though, acquisition leaders mitigate and manage these risks through training, policies, and procedures. This same approach applies to the use of AI in acquisition. As occurred when the internet and computers became tools to assist in acquisition and contract management, use of this new tool in making decisions within the acquisition and contracting process carries similar, if not exponential, national security, financial, and legal implications.

Procedural Risks – If an acquisition office plans to use AI in a source selection or acquisition decision-making, the process must be able to withstand significant legal scrutiny. Any decision that the AI system supports must be substantiated by evidence of the reasoning underlying the decision. Furthermore, as protest decisions are released, AI systems must adapt to incorporate the most recent case law/guidance established by the GAO or the U.S. Courts to supersede any previous guidance. With this understanding, acquisition offices should not use an AI system as the sole decision-maker, but rather as a decision support tool to enhance the knowledge of the human decision-maker. The AI-based system should be considered as a verification tool or an advisor. Acquisition programs must consider additional risks and must proceed cautiously as the Federal Government adopts strategies for implementing AI systems in support of their acquisition processes.

Ethical Dilemmas – With the prevalence of socio-economic programs that ensure small and disadvantaged businesses receive adequate and direct support in the U.S. Government contracting system, training AI models to interpret and weigh the importance of the business type in making an informed and accurate decision presents a unique challenge. For this reason, keeping a "human in the loop" is essential to the introduction and operation of AI in acquisition. All decisions must be fully transparent and traceable. Further, the use of proprietary information to train the model may create significant privacy concerns. For an AI-assisted acquisition, the models must be thoroughly reviewed, vetted, and quarantined to prevent unintended influence.

AI Bias – AI bias refers to the presence of systematic errors in the outputs generated by ML algorithms which can arise from various sources, including the data used to train the algorithms, the design of the algorithms themselves, or the objectives set by the developers (Manyika, 2022). The introduction of biased information into the corpus (the book of knowledge) or into the training data can undermine the benefits of AI-based technologies, leading to outcomes that compromise the integrity and fairness of a procurement effort. Discrimination caused by AI bias can lead to favoritism towards certain solutions based on irrelevant factors, ultimately reducing competition and increasing costs to the Government. Furthermore, AI bias can cause inaccurate predictions or evaluations, resulting in suboptimal decision-making within the acquisition process. Programs should thoroughly vet both the algorithms and training for bias. Furthermore, following the completion of the task or project, the model should be wiped clean of all data or destroyed. This will prevent unintentional exposure of trade secrets, source selection information, proprietary information, and private information.

Malicious Intent – Knowingly or unknowingly manipulating data to influence an acquisition would be extremely effective and dangerous. Influence operations conducted by adversaries consist of hundreds of thousands of social media profiles with similar hashtags, phrases, or reshared topics. Thus, it would be easy for an adversary, competitor, or interested party to skew research, and possibly an associated AI model or tool, in one direction by corrupting the data being ingested by removing or adding information. With the AI models running on algorithms that generate inferences based upon the data to which they have been exposed, the risk of so-called ‘hallucinations’ arises. Hallucinations in data can be extremely dangerous because they are not based on facts and therefore are simply untrue and inaccurate. Ensuring transparency within the decision-making process means that as training, analysis, and decisions are published and become accessible, the attack landscape becomes larger. More access by nefarious actors creates a larger attack surface and better understanding of how to manipulate, obfuscate, or obtain unauthorized access.

Deep Fakes – AI and deep learning techniques have enabled the creation of exceptionally realistic manipulated digital content, often in the form of images, videos, or audio. This phenomenon, known as “deep fakes,” signifies the application of advanced AI techniques to producing counterfeit content (Taha et al., 2022). Candidate vendors can use deep fakes to fabricate credentials, such as documents, images, or videos, falsely showcasing a team’s expertise, experience, or qualifications, and thus making their proposal appear more attractive to the Government. Vendors competing for the contract, or third parties, could use synthetic media to conduct a virtual presentation, answer questions, and perform a variety of demonstrations. As a result, Government agencies must exercise increased vigilance in verifying the authenticity of information presented in proposals. They should conduct thorough background checks on contractors’ credentials and experiences, strictly verify financial statements, and utilize software or tools to detect deep fake images, videos, or audio.

AI Poisoning – AI poisoning is a type of cyberattack conducted by injecting malicious code or misleading data into the AI training code or dataset. Using AI systems to facilitate Federal Acquisition processes without mitigation control for AI poisoning could potentially lead to biased decision-making, manipulated activities, delays in procurement, and mistrust in systems. To address the risks of AI poisoning in the Federal Acquisition process, it is essential that programs implement a robust process for data validation and verification, following the principles of “Trust but Verify,” and monitor malicious activities in Government AI systems.

Cyber-Enabled Espionage – As AI is integrated into the acquisition process through the use of code, software, and the internet, the attack surface to which these vectors can be exploited increases through cyber-enabled espionage. Obtaining source selection information may give a competitor an illegal advantage or an investor a lead, but, more significantly, adversaries constantly seek to undermine U.S. security systems to obtain designs, intellectual property, or trade secrets at no cost after a company has spent millions of dollars on research and development (Federal Bureau of Investigation, 2022). Writing a spear phishing attack and coding a malicious email may have previously taken one person 3 hours; AI can ingest a person’s biography to create and send an email with malicious code within seconds. This increases the risk that Federal Acquisition staff will encounter exploitation of proprietary, source selection, and national security information.

Mitigating the Threats, Risks, and Biases in AI

While the threats detailed in the previous section may trigger a very reasonable "extreme caution signal" for the risk-averse, acquisition programs can, and should, employ various strategies to mitigate these risks. The National Institute of Standards and Technology (NIST) is developing Federal AI acquisition testing and evaluation capabilities and guidance for other Federal agencies to reference NIST's Artificial Intelligence Risk Management Framework. Within its AI Risk Management Framework (AI RMF 1.0), NIST identified the "characteristics of trustworthy AI systems" as: Valid and Reliable, Safe, Secure and Resilient, Accountable and Transparent, Explainable and Interpretable, Privacy Enhanced, and Fair – with Harmful Bias Managed (NIST, 2024).

Ensuring that each program considers these characteristics as criteria that determine acceptability of any AI system used in Federal Acquisition will be paramount in ensuring the success of the system(s) and optimizing Federal Acquisition processes with AI. The list below summarizes some of the best practices for inspecting and approving AI-based systems. These guidelines provide a framework that helps Federal agencies to ensure their AI systems are not only efficient and effective, but also ethical and in line with necessary Federal regulations.

- **Clearly Defined Requirements:** Use non-ambiguous, precise and succinct language when defining system requirements to limit the risk of misinterpretation and misunderstanding. By clearly defining requirements, programs will ensure that potential vendors have a fundamental understanding of the intent and outcomes the system should strive to achieve.
- **Vendor Assessment:** Evaluate the credibility and previous performance of AI application vendors, especially their performance history with AI deployments. Acquisition professionals can assess the vendors' approach from an ethical and compliance perspective based on current standards, regulations, and best practices.
- **Transparency and Comprehensibility:** Ensure the AI system's decision-making processes are transparent and explainable. This is essential to understanding how the AI system arrives at conclusions and ensuring accountability.
- **Ethical Considerations:** Incorporate ethical guidelines in the acquisition process. This includes considerations pertaining to privacy, data security, fairness, and avoidance of bias.
- **Testing and Validation:** Implement rigorous testing and validation protocols to test the AI system in controlled environments and validate its performance against predefined criteria.
- **Third-Party Audits:** Assess the AI system's compliance with standards and absence of biases through independent audits or certifications from reputable third parties. This will also act as an independent check to mitigate any unintentional biases in an organization's culture or processes.

Federal Acquisition professionals may find themselves leveraging AI to augment or expedite acquisition processes for the procurement of an AI system. In such cases, both sides of the acquisition will present threats and risks, which demand careful attention and oversight of the systems, their logic processes, and their auditing outcomes to ensure that the mitigation strategies are effective against the threats previously described as well as unknown threats yet to emerge.

Recommendations

First and foremost, leaders in the acquisition field must accept the inevitability of AI permeating every facet of the acquisition lifecycle. Like every technical revolution that has preceded it, this technology cannot be wished away. Defense policy makers and acquisition professionals who ignore the promise of AI squander an incredible opportunity while inevitably harming the national security of the United States. Rather than adopting a wait-and-see philosophy or adopting a philosophy of ignoring AI, acquisition leaders must embrace AI in a safe and responsible way. Doing so will open the door to a new acquisition revolution, enabling acquisition teams to procure higher quality capabilities and solutions at lower cost and at a much greater speed.

The natural question is, "what next?"

1. Recommended Next Steps

Do not categorically deny the use of AI in Acquisition

Acquisition leaders must understand the potential that AI has in enhancing the outcomes of the acquisition process. Unrestrained fear of this technology is undeniably a formula for deterioration in defense and the efficient means of executing the Agency's missions. Rather, as technologists, acquisition leaders must embrace their role on the front lines of technology adoption and implementation.

Across the acquisition lifecycle, acquisition leaders must identify opportunities to implement AI, use AI to enhance their organizational outcomes, and mitigate the risks that come along with this technology. Education, communication, and prototyping are paramount. Educate acquisition teams, understand the risks, leverage institutional systems for risk management and security, and above all, relentlessly pursue solutions that responsibly harness this power.

Establish a cross-functional team of experts for implementation and oversight, and ensure AI transformational leadership at various levels

Gaining momentum for organization-wide adoption and implementation of AI systems with a top-down strategic focus on executing discrete and shared mission goals and objectives requires an AI Governance Board composed of the various organizational stakeholders. This board should focus on implementing AI at the speed of relevance. AI adoption also requires a dedicated group of professionals who maintain current awareness of AI development, deployment, and maintenance trends and best practices to advise the Governance Board when actionable insights become necessary.

2. AI Governance

Regardless of an organization's size or mission, any governance approach must address organization-wide guidance for establishing an approach for acquiring and implementing AI systems, while respecting the nuanced needs of various components within the organization.

Patrick T. Blitgen (2024), PhD, author of *AI for Defense and Intelligence*, offers four strategies that organizations can leverage to begin AI governance but must be tailored to their need and intended use of AI.

Establish clear policies: Policies should be in place to guide the ethical use of AI, the handling of data, and the lifecycle management of AI models. These policies should be regularly reviewed and updated to keep pace with technological advancement.

Enhance transparency: Organizations should strive for transparency in their AI systems, including clear documentation of data sources, model architectures, training procedures, and decision-making processes.

Implement robust oversight mechanisms: This includes establishing dedicated governance bodies or committees, conducting regular audits and reviews, and implementing mechanisms for reporting and addressing issues.

Provide training and education: All stakeholders, from decision makers to end-users, should be educated about the principles and practices of AI governance. This will help ensure that everyone understands their roles and responsibilities and can make informed decisions about using AI.

3. Task Force

For an AI Governance Board to be effective, it must focus on taking inventory of the organization's current AI capabilities and potential areas of application to responsibly and strategically monitor how to mature the implementation approach. However, with summits on AI advancement occurring almost continuously – at the time of publication (January 2024), summits have spanned the global level (World Economic Forum) and individual Federal entity levels such as the Federal Communications Commission – the Governance Board cannot extract relevant and impactful updates from the flood of daily developments.

The DoD's (2023) AI Hierarchy of Needs (pictured below) serves as a frame of reference to distinguish between the AI Governance Board's and its Task Force's areas of responsibility. The Governance Board would be responsible for setting the foundation of quality data, subsequently building the pyramid upward, and then coordinating internal efforts by the various components charged with continuously implementing enabling actions. The Task Force would be responsible for maintaining situational awareness across the public, private, academic, and international community of external factors that could affect different levels of the pyramid. This Task Force would not function merely as a watch dog but would (1) brief the Governance Board on impact areas and (2) suggest courses of action. This would require an interdisciplinary group of AI SMEs with policy, technical, and behavioral backgrounds.

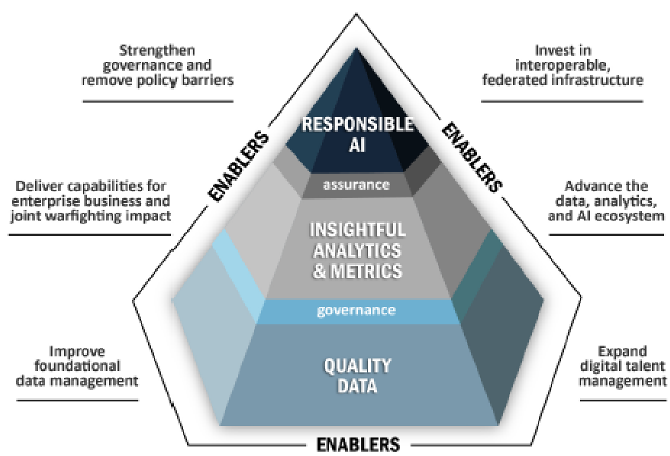


Figure 1.

Establish a holistic AI acquisition framework

Government agencies must create a comprehensive framework encompassing the complete lifecycle of AI systems from inception to implementation and ongoing maintenance. This applies to both commercial off-the-shelf (e.g., IBM Watson, ChatGPT) and in-house-developed AI products and services (MITRE, 2023) and is in line with the National Security Commission on Artificial Intelligence's (NSCAI's) final report (Schmidt et al., n.d.). The framework should incorporate guidelines for

A. <u>AI system requirements definition:</u> Clearly define performance, safety, security, and ethical requirements for AI systems.
B. <u>AI system design and development:</u> Establish best practices for designing and developing AI systems that meet the defined requirements.

C. AI system testing and evaluation: Develop rigorous testing and evaluation methodologies to ensure AI systems perform as intended and meet safety, security, and ethical standards, in line with the NSCAI's suggestions.
D. AI system deployment and monitoring: Provide guidance on deploying AI systems in operational environments and monitoring their performance, safety, and security.
E. AI system maintenance: Establish processes for maintaining and updating AI systems to ensure their continued safety, security, and trustworthiness.
F. AI maturity model: Build a model of AI integration and maturation within the organization according to its performance goals, mission objectives, and organizational capacity.

Promote transparency and accountability both internally and with external partners

Government organizations can achieve transparency in AI system development and deployment by providing clear documentation of system capabilities and design, objectives, limitations, and potential risks. This must go beyond technical documentation and explain in plain English where in the process AI is being leveraged so that humans either using the AI's output or working with it (i.e., in the loop) can justify and explain the final product/decision. The Government should prioritize safety and ethics to ensure that AI development follows Federal ethical guidelines and safety principles to minimize potential harm to the public and promote safe and effective AI systems that safeguard civil rights. This will enable Government stakeholders to make informed decisions about AI system acquisition and deployment, while mitigating the likelihood that risks will materialize (OSTP, 2023).

Foster collaboration and information sharing

Organizations should consider beginning with small, specialized AI systems to ease the transition into broader AI use and conduct outreach to other Government agencies, industry partners, and academia leading the charge on AI in acquisition efforts. They should use internal and external lessons learned to shape and institutionalize best practices. This aims at enhancing the Federal Acquisition process and promoting the development and adoption of best practices for AI system safety, security, and trustworthiness (OSTP, 2023).

Similarly, Government organizations should leverage proven acquisition techniques to spur a competitive atmosphere and accelerate innovation and acquisition. For example, challenge-based events are known to garner participation from all parts of the private sector (large businesses, small businesses, and academia). High-profile problem spaces, such as the Defense Advanced Research Project Agency's (DARPA's) Grand Challenge, accelerated advancement in autonomous vehicle technologies and incentivized involvement through a $1 million prize, the winner being a Stanford University team. CDAO is currently undertaking a similar effort, dubbed an AI Bias Bounty, to increase the speed at which the Government acquires leading-edge technology (DoD, n.d.). CDAO could consider elevating this challenge framework to the intergovernmental level by showcasing AI tools that may offer cross-cutting benefits and then sponsoring private sector challenges to meet shared problems in the Federal arena and thus accelerate government-wide access to trusted solutions.

Advocate policy and regulatory updates

Government agencies should regularly evaluate and update their AI systems to ensure they continue to meet the evolving needs of the Federal Acquisition process. By adopting AI-driven solutions, the Government can significantly improve the efficiency, effectiveness, and security of its acquisition processes, ultimately delivering better value for taxpayers and enhancing the overall quality of public services (Hamilton, 2022).

Acquisition leaders are the first line of defense to understanding the gaps and needs in this space. As situations identify various policy and regulatory shortcomings, acquisition leaders must use their platform to communicate and educate their findings.

Understand the Drivers of AI Success

Pilot projects and ideation sessions are critical first steps in AI adoption to identify impactful use areas and obtain organizational user buy-in. After identifying tasks for AI integration, organizations should draw on lessons from current AI adoptions, anticipating challenges and seizing opportunities.

PREREQUISITE: Prioritize data hygiene across datasets. Well-maintained data is crucial in developing effective AI tools and preventing errors or hallucinations.

- **Data Management in AI:** Effective data management is pivotal in AI implementation, serving as the foundation for accurate and reliable AI outputs. High-quality data ensures the AI systems are trained on relevant and comprehensive information to deliver desired outputs, reducing the risk of biases or errors. This has particular importance in environments where AI influences decisions that have significant consequences, such as developing requirements and determining vendor viability. As AI continues to evolve, the role of robust data management becomes even more critical, ensuring AI tools perform not only optimally but also ethically and responsibly.
- **Organizational Change Perspective on AI Implementation:** AI implementation represents a significant shift in organizational dynamics, necessitating a thoughtful change management approach. It requires balancing technological advancements with human factors, ensuring that employees are adequately prepared and supported. This transition involves not just the adoption of new technologies but also a cultural shift towards embracing digital transformation. Organizations successful in AI integration often prioritize continuous learning, adaptability, and employee engagement. Organizations must recognize that the human element is as crucial as the technology and that this approach helps to ensure a smooth and effective integration of AI into existing workflows.
- **Contrasting Culture and AI Adoption:** Organizational cultures resistant to change or lacking technological literacy face greater challenges in integrating AI into their processes. These dynamics highlight the necessity of building a culture that not only embraces technological advancements but also actively prepares for them through training, awareness, and leadership support. The success of AI adoption hinges not just on the technology itself, but on how well it aligns with and is nurtured by the organization’s cultural ethos.
- **Establish a maturity model:** Immediately establish a maturity model, overseen by the organization’s appropriate governance board, such as the Department of Energy’s AI and Technology Office (AITO) and the VA’s NAII. This model should guide the transition of AI tools from development to implementation and be aligned with workforce and mission needs. Creating such a model also presents an opportunity to distinguish between AI-generated automation and general software automation, avoiding potential misunderstandings.
- **Identify the tasks that AI could absorb or assist in:** Determine the types of tasks best suited to AI integration, focusing initially on “low-hanging fruit” tasks with low complexity/risk. This approach would allow documentation of organization-specific lessons before progressing to more complex AI-compatible tasks. For example, SMEs suggested using AI to automate the drafting of iterative reports, enabling human experts to focus on authoring initial reports and reviewing subsequent versions.

Challenges: Be aware of these, but they can be overcome by the guidance provided!

- **Acquisition professionals largely lack technical backgrounds:** Since many acquisition professionals are unfamiliar with technology, they require training in AI usage. Developing or hiring in-house AI expertise takes time, so users primarily need to learn how to use AI tools responsibly, confidently, and effectively. Industry adopters of AI who struggled to promote end user adoption, specifically of chatbots, reported, "most of us are using the technology in a suboptimal way, largely because the tech companies gave us poor directions" ("We're Using AI Chatbots Wrong. Here's How to Direct Them," 2023). Users need a clear and direct interface (i.e., prompts) to explain how to pose the right questions to elicit accurately sourced responses.
- **Misconceptions create pushback:** Programs should directly address misconceptions as they arise. It is crucial to emphasize that humans will remain integral in decision-making. AI will enhance, not replace, human judgement. Security is paramount. All AI integrations will occur in secure, Government-approved environments, with necessary security protocols and application programming interfaces (APIs) in place. The bottom line is that programs will not integrate AI into any of their decision-making processes without the appropriate security assurances.
- **Decisions must be traceable:** At critical points, programs require comprehensive documentation to trace how AI analysis informs decisions. For instance, AI could identify the source(s) of its findings, such as the FAR (and the specific section and sub-section) or vendor documents (and specific page and paragraph numbers).

Develop training and education programs for an 'AI Ready Workforce'

Prior to preparing the Federal workforce to use AI, the organization must understand its employees' experience and how changes will affect them. This will help to inform communication and education campaigns. *"Part of the challenge is that AI is evolving so quickly that frameworks, tools, and guidance will need to be continuously updated and improved as we learn more"* (AI COE, n.d.).

Organizations must foster a cultural shift towards embracing curiosity and encouraging staff members to ask questions about AI early, often, and repeatedly. The GSA's AI Guide for Government offers these organizational suggestions to foster responsible and trustworthy AI (AI COE, n.d.):

- **Focus on the root problem.** *Why is AI being considered as a solution? Is it the best option to solve this problem?*
- **Be accountable to the users.** *Establish clear roles and responsibilities and ensure the outcomes of the systems are justifiable to the users who interact with it.*
- **Define and avoid harm.** *Evaluate what possible harms could be and how bias might cause disparate, negative impacts to create mitigation strategies to reduce that possibility.*
- **Monitor the outcomes.** *Are there regular management reviews of changes? Are the systems auditable so that the drivers of incorrect or inequitable outcomes can be identified and fixed?*

Once an organization has established this culture of continuous learning, it must turn its attention to the individuals who comprise its workforce.

First and foremost, acquisition programs must "avoid centralizing AI practitioners and leaders in one unit. AI talent must be accountable to the business needs and therefore should exist across the organization" (AI COE, n.d.). This talent will constitute the bedrock of knowledge to support upskilling of their peers. The AI Guide for Government also recommends that this AI talent "be involved in further talent recruitment, certification, training, and career path development for AI jobs and roles" (AI COE, n.d.).

4. Continuous Learning and Upskilling

The existing workforce must understand the AI system's decision-making process so that they can trust its output and provide "explainability" and traceability as they use it to inform their next steps. Developing and retaining AI talent among new staff members will require (AI COE, n.d.):

- Incentives for skill development
- Formal education opportunities
- Optional training, conferences, and exchanges with industry and academia.

The workforce will be learning about a technology that is itself still rapidly evolving; therefore, these support resources should be designed to encourage personnel to provide feedback on ineffective or irrelevant processes.

Adopt a Security-Forward Mindset

To ensure the Federal Acquisition process can fully utilize, operationalize, and conceptualize the use of AI, acquisition offices must ensure they understand the severe consequences of not utilizing AI/ML properly. They must consider procedural risks, ethical dilemmas, malicious intent, deep fakes, AI poisoning, AI bias, and cyber-enabled espionage when adopting this novel technology, which has the potential to change the future of acquisition. With technology, change, and tax dollars come increased scrutiny, oversight, and importance. If acquisition offices do not understand, consider, and address the issues above when utilizing a radically new capability, the public and Congress may have reduced confidence in the Government's ability to fairly and properly manage taxpayer funds and deliver essential Government functions.

Conclusion

The defense acquisition system rarely gets an opportunity to implement the solutions that it acquires for the warfighter to improve its own ends. AI offers a compelling case for acquisition leaders seeking to enhance their impact on the defense ecosystem. While there are many different use cases for AI systems in the acquisition process workflow, the underlying assumption across the acquisition enterprise must be that these technologies are ubiquitous and essential to mission needs/objectives. As leaders on the front lines of technology adoption, acquisition teams must drive a risk-informed policy of seeking, understanding, and employing these vital technologies. Ultimately, the successful adoption of AI is dependent upon seeing the promise of AI and safely bringing it into the acquisition lifecycle.

References

AI COE. (n.d.). *AI guide for government*. https://coe.gsa.gov/coe/ai-guide-for-government/print-all/index.html

AI inventory — tech at GSA. (n.d.).

Blitgen, P. T. (2024). *AI for defense and intelligence.*

Brynjolfsson, E., Li, D., & Raymond, L. (2023). *Generative AI at work*. https://doi.org/10.3386/w31161

Congressional Research Service. (2023, August 4). *Artificial intelligence: Overview, recent advances, and considerations for the 118th Congress* (CRS Report No. R47644). https://crsreports.congress.gov/product/pdf/R/R47644

DelTek. (2023, August). *Federal artificial intelligence landscape, 2024* [Slide show]. GovWin.

Department of Labor. (n.d.). *Artificial intelligence use case inventory.*

Department of State. (2023). *AI inventory – United States*.

DoD. (n.d.). *CDAO launches first DOD AI bias bounty focused on unknown risks in LLM*. https://www.defense.gov/News/Releases/Release/Article/3659519/cdao-launches-first-dod-ai-bias-bounty-focused-on-unknown-risks-in-llms/

DoD. (2023, June 27). *2023 data, analytics, and artificial intelligence adoption strategy*. https://media.defense.gov/2023/Nov/02/2003333300/-1/-1/1/DOD_DATA_ANALYTICS_AI_ADOPTION_STRATEGY.PDF

Federal Bureau of Investigation. (2022, July 21). *Combating economic espionage and trade secret theft*. https://www.fbi.gov/news/testimony/combating-economic-espionage-and-trade-secret-theft

Federal Register. (2020, December 8). *Promoting the use of trustworthy artificial intelligence in the federal government*. https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government

Gill, J. (2023, November 7). *DISA wants to use AI as a 'digital concierge' for its workforce*. Breaking Defense. https://breakingdefense.com/2023/11/disa-wants-to-use-ai-as-a-digital-concierge-for-its-workforce/?utm_campaign=BD%20Daily&utm_medium=email&_hsmi=281636445&_hsenc=p2ANqtz-8NEr73yhYBheB8kT5ouCfqiPUcGVcKOsvLrqgFHUf-c0Q-ZsaquzJsNs0gEOKYEB2eZL9rRljaieEe8xAwKQ7RmNAMXQ&utm_content=281636445&utm_source=hs_email

Hamilton, T. (2022). Artificial intelligence acquisition guidebook. In *Department of the Air Force / MIT Artificial Intelligence Accelerator*. https://aia.mit.edu/wp-content/uploads/2022/02/AI-Acquisition-Guidebook_CAO-14-Feb-2022.pdf

Harvard University Information Technology. (n.d.). *Generative artificial intelligence (AI)*. https://huit.harvard.edu/ai

Heckman & Heckman. (2023).

HHS. (2023). *2023 public inventory of AI use cases – HHS.gov*

IBM. (n.d.). *What is natural language processing?* https://www.ibm.com/topics/natural-language-processing

Kanowitz. (2023).

Manyika, J. (2022, November 17). What do we do about the biases in AI? *Harvard Business Review*. https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai

MIT Sloan. (2021, April 21). *Machine learning, explained.* https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

MITRE. (2023). *A sensible regulatory framework for AI security*. https://www.mitre.org/sites/default/files/2023-06/PR-23-1943-A-Sensible-Regulatory-Framework-For-AI-Security_0.pdf

Newman, D. (2024, January 25). AI rising: 7 technology leaders provide a view from Davos. *Forbes*. https://www.forbes.com/sites/danielnewman/2024/01/24/ai-rising-7-technology-leaders-provide-a-view-from-davos/amp/

NIST. (2024, January 5). *AI risk management framework*. https://www.nist.gov/itl/ai-risk-management-framework

OSTP. (2023, November 22). *Blueprint for an AI bill of rights.* The White House. https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Rich, B. (2020, October 27). How AI is changing contracts. *Harvard Business Review*. https://hbr.org/2018/02/how-ai-is-changing-contracts

Schmidt, E., Work, R., Catz, S., Horvitz, E., Chien, S., Jassy, A. … McFarland, K. (n.d.). *Final report: The National Security Commission on Artificial Intelligence*. https://cybercemetery.unt.edu/nscai/20211005220330/https://www.nscai.gov/

Taha, M., Khudhair, W. M., Khudhur, A. M., Mahmood, O. A., Hammadi, Y. I., Al-Husseinawi, R. S. A., & Aziz, A. (2022). Emerging threat of deep fake: How to identify and prevent it. *International Conference on Future Networks & Distributed Systems*, *6*. https://doi.org/10.1145/3584202.3584300

U.S. Department of Treasury. (2024). *Artificial intelligence (AI) use cases*

USDA. (n.d.). *2023 public inventory of AI use cases - USDA.csv.*

We're using AI chatbots wrong. Here's how to direct them. (2023, July 20). *The New York Times*. https://www.nytimes.com/2023/07/20/technology/personaltech/ai-chatgpt-bing-directions.html?name=styln-artificial-intelligence®ion=TOP_BANNER&block=storyline_menu_recirc&action=click&pgtype=Article&variant=undefined

The White House. (n.d.). *Government use of AI*. https://ai.gov/ai-use-cases/

The White House. (2023a, January 24). *National artificial intelligence research resource task force releases final report*. https://www.whitehouse.gov/ostp/news-updates/2023/01/24/national-artificial-intelligence-research-resource-task-force-releases-final-report/

The White House. (2023b, May 23). *Fact sheet: Biden-Harris administration takes new steps to advance responsible artificial intelligence research, development, and deployment*. https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/23/fact-sheet-biden-harris-administration-takes-new-steps-to-advance-responsible-artificial-intelligence-research-development-and-deployment/

The White House. (2023c, October 30). *Executive order on the safe, secure, and trustworthy development and use of artificial intelligence*. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

Wolf Stake. (n.d.). *GPTs for use*. https://www.wolfstake.org/home/gpts-for-use

Young, S. D. (2023). *Proposed memorandum for the heads of executive departments and agencies*. Executive Office of the President. https://www.whitehouse.gov/wp-content/uploads/2023/11/AI-in-Government-Memo-draft-for-public-review.pdf

Improved Forecasting of Defense Acquisition Program Performance Using Digital Twin Models of a Revenue Centric Versus Cost Centered Approach (Enhanced Earned Value Management (E2VM))

COL (Ret., USA) Raymond Jones—is the Chair and Professor of Practice in Defense Program Management in the Department of Defense Management at the Naval Postgraduate School (NPS). Prior to NPS, he was the CEO and co-founder of a strategic consulting company which focused on helping high technology firms improve their innovation strategies. He served as the Deputy Program Executive Officer for the Joint Tactical Radio System (JTRS) and managed three major programs in the DoD. In addition to multiple operational assignments in combat aviation units, he is a 1995 graduate of the U.S. Naval Test Pilot School and a 1983 graduate of the United States Military Academy. He has a Bachelor of Science degree in Aerospace Engineering, a Master of Science degree in Aeronautical Engineering, and a master's in business administration, and is a distinguished graduate with a master's degree in National Resource Strategy from the Industrial College of the Armed Forces.

Abstract

This paper represents a new approach to defense acquisition program forecasting during the development phase of the program life cycle. It will be the first of three research papers that will attempt to improve insight into how a program performs and will offer a method by which future programs offices will be able to simulate their program before beginning in order to develop an optimal acquisition strategy. Specifically, the purpose of this research is to explore if a digital twin of the defense acquisition development phase of an acquisition program of record can enhance a program manager's decision-making ability by revealing unforeseen patterns in program behavior. Additionally, this research will demonstrate a new way of measuring value and return on investment of a defense program of record, to provide decision-makers with an alternative to traditional methods of decision-making that are based upon budget and cost comparisons. This paper represents the initial research proposal and method, which will set the conditions for exploration of new theory and methods of how complex programs could be planned and executed in the future.

Introduction

Background

The Defense Acquisition process is complex and challenging and involves the acquisition of goods and services for the Department of Defense (DoD). The foundation of the DoD Acquisition process is to generate knowledge that is relevant to the development of operationally relevant goods and services. It is fundamentally a risk reduction process that follows a prescribed life cycle that transitions technological insight and development from a low level of maturity to a level sufficient to transition to production and delivery to the customer. The acquisition process involves numerous steps, including research and development, testing and evaluation, production, and sustainment. Despite the importance of the defense acquisition process, it faces several challenges that can impact its effectiveness and efficiency. One of the most significant challenges in the defense acquisition process is transitioning critical technology from a technology base to a program of record organization that has the authorities and resources to develop technologies into viable operational capability. This transition is commonly referred to crossing the "valley of death."

The "valley of death" is a term used to describe the gap between research and development and the commercialization of new technologies. This gap can be particularly challenging in the defense acquisition industry, where the development of new technologies is critical to the success of military operations. Developing new technologies can be expensive, and there is no guarantee that the technology will be successful. This can make it difficult for companies to secure funding for research and development, which can slow the pace of innovation.

The acquisition process can also take years to complete, and it involves numerous stakeholders, including government agencies, contractors, and military services. This lengthy process can result in delays, cost overruns, and the delivery of outdated technology. Crossing the valley of death in defense acquisition is a complex and challenging problem that requires collaboration between stakeholders, increased funding for research and development, and a streamlined acquisition process. Addressing these challenges will require innovative solutions and a willingness to work together to ensure that new technologies are developed and deployed to support military operations.

DoD programs typically follow a prescribed path in which technology solutions are vetted through a complex series of administrative processes. Once budgets are set and vendors are identified, it is extremely difficult for a program manager to stray from the prescribed acquisition strategy, reflecting little agility to adapt to program volatility and integrate new innovative solutions outside of the predetermined acquisition baseline. Additionally, current forecasting models in the management of defense capability solutions provide the program manager limited insight into whether current planning will be successful as the program transitions through the Defense Management program life cycle.

Making decisions in this highly complex and restrictive environment requires new ways of understanding the data and information flowing through the development life-cycle process. Decision-makers need to better understand how knowledge is created and how to make choices among an endless set of options. Knowledge-making is the process of creating new knowledge or insights through research, analysis, and critical thinking. It involves the synthesis of existing knowledge, the identification of gaps in knowledge, and the development of new ideas and perspectives (Choo & Bontis, 2002). The creation of knowledge is useful for improved decision-making before a program under development encounters critical problems. Information networks facilitate the flow of knowledge and communication between individuals and groups, creating the opportunity for choices and feedback within the development life cycle of the defense acquisition process.

The defense acquisition framework can be viewed as an information network with stakeholders as nodes and communication channels as edges. The stakeholders in the network include the DoD, contractors, suppliers, and other entities involved in the acquisition process. Viewing the defense acquisition process as an information network allows stakeholders to identify bottlenecks and inefficiencies in the process. By analyzing the flow of information and communication between discrete events in the process, stakeholders can identify areas where communication paths are overloaded and become inefficient relative to the planned strategy (Scott et al., 2013). Specifically, Scott et al. (2013) address the challenges in which multiple stakeholders allocate tasks for multiple decision criteria while attempting to satisfy multiple groups, confounding the decision-making process. Viewing the defense acquisition process as a complex network of information flow that also represents increasing value in the form of work, requires consideration of multiple conflicting criteria and the consideration of uncertainty.

The theory of information networks in an organization is based on the concept of social network theory and involves understanding of how individuals and groups are connected through communication channels and information flows (Wasserman & Faust, 1994). Additionally, information networks provide a framework by which a more in-depth understanding of decision-making can be understood. This is the foundation upon which this research derives its critical importance. By developing a virtual replica of the defense acquisition environment and the integration and application of decision theory and tools, it may be possible to forecast program management performance of a program as it moves through the development phase of the defense acquisition life cycle.

The development phase in the defense acquisition life cycle is a period in which the designed system or capability is transformed into a tangible and operational product. Critical decisions are made during this phase that influence how well the system is being managed and whether the process is being optimized to provide as much value as possible in results of the management decisions being made. This phase follows the initial stages of concept exploration and technology development and precedes the production and deployment phases. The development phase involves a series of activities, milestones, and assessments to ensure that the system meets the established requirements defined by the user community and is ready for subsequent production.

Not since Frederick Taylor has there been any significant change in how management and more specifically program management change the way programs are planned and executed. Taylor's work set the conditions upon which modern day development and production programs are structured and conducted. Taylor established standardization and process in the program management field that are still in use today (Taylor, 1911). While these practices provide a good foundation for planning and monitoring a program's process relative to a baseline, they do not provide a complete insight into why programs tend to go astray. In effect, modern day practices are still regressive in nature and provide little opportunity for informed prediction. There is a distinct lack of clarity in program performance in the vast sea of data and information regarding how programs are performing and more importantly, how they will perform in the future. New methods and metrics need to be discovered to begin to address this challenge.

Research Problem

The problem is that the DoD does not have a reliable and measurable forecasting process with which to determine whether capital investment in programs under development is being optimized to produce the desired output within program performance constraints. The DoD focuses on traditional cost theory as a principal driver for assessing program performance, which is exacerbated by historically rigid accounting processes. Current methods rely on the assumption that program cost and schedule can be estimated based upon predetermined frameworks (GAO, 2009). These frameworks rely on historical performance of programs as well as parametric models that presume to predict the complexity of a program under development with regard to its impact on cost and schedule. There is currently no capability that allows decision-makers to simulate future performance of a program under consideration, resulting in best effort analysis based upon subject matter expertise and past performance. Since the DoD has no way to simulate a program of record prior to its inception, using cost predictions relative to expected performance is the best method currently available to program managers.

The purpose of this research is to explore if a digital twin of the defense acquisition development phase of an acquisition program of record can enhance a program manager's decision-making ability by revealing unforeseen patterns in program behavior. A digital twin is a virtual representation of an object, system, or process that spans its life cycle. Digital twins are often used to simulate the "real world," using machine learning, artificial intelligence, and reasoning to improve decision-making (IBM, 2023).

This leads us to the following hypothesis:

H1: A virtual replica with an integrated artificial intelligence tool has the potential to improve decision-making leading to improved efficiency and cost effectiveness of defense acquisition programs during the development phase of the life cycle.

Using a digital twin of the defense acquisition process offers the potential to provide valuable insight and knowledge that can help improve the acquisition process. For example, a virtual replica of the defense acquisition process can reveal patterns of behavior and information exchange that will allow decision-makers to test different scenarios and strategies without impacting the physical object or system. This will help decision-makers understand the potential outcomes of different decisions allowing for more informed choices. Identification of bottlenecks and inefficiencies will help stakeholders to optimize processes, reduce costs, and improve efficiency. Additionally, patterns not well understood in process documents will facilitate a better understanding of risks and opportunities more effectively.

A virtual replica of the defense acquisition process can simulate the impact of changes in the process and will provide a simulation environment for the actual environment that will reveal potential choke points and alternate paths and courses of action for the decision-maker. This can help stakeholders to understand the potential impact of changes before they are implemented, reducing the risk of unintended consequences and improve decision-making potential. This leads us to our second hypothesis:

H2: A virtual replica with will provide more certainty in a defense program under development and reduce overall risk in program development, thereby increasing the probability that a program will more closely meet its cost and schedule objectives.

Using a virtual replica of the defense acquisition process can provide valuable insights and knowledge that can help improve the acquisition process. Improved decision-making, identification of bottlenecks and inefficiencies, increased collaboration, improved risk management, and a better understanding of the impact of changes are just some of the knowledge that can be gained by using a digital twin of the defense acquisition process.

Research Objectives

There are three primary research objectives for this dissertation.

1. Develop a conceptual framework for integrating virtual replica technology into the defense acquisition life cycle.
2. Investigate the potential of virtual replicas to improve decision-making in the defense acquisition process.
3. Analyze the impact of virtual replicas on decision-making risk assessment and overall efficiency in defense acquisition.

Significance of the Study

The expected outcomes of this dissertation are to provide a comprehensive understanding of the use of digital twins in a Defense Acquisition Program of Record. The study aims to identify the potential benefits of using a digital twin in this context, including improved decision-making, increased efficiency, reduced costs, and the impact of varying the current acquisition pathways to insert critical technologies after a POR is started. This study is important because there is currently no useful simulation environment in the defense acquisition process that allows program managers to virtually execute their programs and observe them under conditions of ambiguity and change. By allowing the PM to simulate program volatility before a program has even begun, decisions can be made prior to program disruption improving the probability of program success. This study will contribute to the fields of decision-making, program management, and socio-technical theory.

Literature Review

This literature review serves as the foundation upon which to extend existing theory in program management. The intersection between decision-making, program management, and information theory is critical in shaping existing paradigm and challenging the status quo, from which little progress has been made regarding new theoretical constructs and knowledge. A detailed study and analysis of the literature across different, yet complementary theories will play a pivotal role in establishing the existing knowledge landscape, identifying gaps, and positioning the research within the broader academic discourse.

The strategy that will be used for this dissertation study will be to organize and categorize the extensive body of knowledge into meaningful categories. The literature will help to address gaps discovered during model development and experimentation and help to contribute to the field of program management and decision-making through new frameworks. To extend knowledge and create new frameworks from which to improve our insight in program management and decision-making, we will need to examine complimentary theories and processes that intersect the dominant field we are studying. While the literature reviewed for this proposal is simply an initial review, it provides the foundation for a more in-depth study of the significant body of knowledge related to this research. I will scope the literature around four key theories that will be augmented with significant literature from the practice of defense acquisition and program management.

The four theories that are most relevant to better understanding the problem at hand are *decision support systems, socio-technical behavior, economic value, and technology trust*. These theories are intertwined in their relevance to being able extend our knowledge on how decisions are made in a complex management process and perhaps offer insight into the path forward to creating decision support and forecasting tools for program management. Technology trust theory is critical in gaining a sense of understanding into the viability of the recommendation and ultimate adoption of new concepts.

Defense Acquisition Programs

We begin with setting the conditions for this research by examining the current defense acquisition system and how the process is currently managed. It is important to fully understand the nature of how development programs are managed and what methods a manager uses to make decisions and attempt to predict future outcomes. This will inform the research strategy and allow us to examine the relevant theories from which to develop the research design.

Defense acquisition programs constitute complex initiatives involving the development, procurement, and management of military capabilities (Fox, 2016; Sullivan, 2017). These programs are frequently characterized by intricacies, uncertainties, and the necessity for effective decision-making throughout their life cycles (Department of Defense [DoD], 2017). Traditional program management approaches often encounter challenges in addressing these complexities, prompting a need for innovative solutions to enhance decision-making processes. The Defense Acquisition process plays a critical role in ensuring that the U.S. Department of Defense (DoD) acquires and delivers effective and efficient weapon systems and capabilities. Predicting performance in this complex and multifaceted process is of paramount importance to avoid costly delays, budget overruns, and failures in delivering the necessary military capabilities.

The defense acquisition environment is characterized by its multifaceted nature, influenced by political, economic, technological, and social factors. Geopolitical considerations, such as regional tensions and international relations, play a pivotal role in shaping defense priorities (Fox, 2016). Additionally, the rapid pace of technological advancements introduces both opportunities and challenges, necessitating constant adaptation in defense acquisition strategies (Elkins, 2019). Budgetary constraints further compound the complexities, requiring efficient resource allocation and strategic decision-making (DoD, 2017) within a complex set of regulatory and statutory constraints.

Stakeholders in the defense acquisition environment encompass a diverse range of actors, including government agencies, defense contractors, military personnel, and the broader public. Government agencies, such as the DoD, formulate policies, set acquisition priorities, and oversee the execution of programs (Elkins, 2019). Defense contractors contribute to the development and production of military systems, fostering collaboration with the government to meet program objectives (Fox, 2016). Military personnel are essential stakeholders, providing input on operational requirements and utilizing the acquired capabilities in the field.

Finally, defense acquisition programs are comprehensive initiatives designed to address the nation's security needs by developing, procuring, and sustaining military capabilities (DoD, 2017). These programs typically follow a structured life cycle that includes requirements definition, system design and development, production, testing, and sustainment (Fox, 2016). The defense acquisition process is governed by regulations and guidelines, such as the Defense Acquisition Guidebook, which outlines best practices and procedures for effective program management (DoD, 2017).

Decision-Making in Program Management

As technology continues to advance, the ability to create digital replicas of both the physical and decision-making environment will allow researchers to accelerate their learning in the virtual space. Digital twins allow for more in-depth analysis and risk-taking in the virtual world allowing the researcher to change the boundary conditions without impacting the real world, thus providing a richer environment for learning and experimentation. Digital twins provide a comprehensive view of the physical system, enabling decision-makers to access real-time data and predictive simulations. Decision-makers can make more informed choices by considering the implications of various alternatives. This aligns with the principles of bounded rationality and the adaptive decision-making models, where individuals aim to achieve the best outcome given their limited cognitive resources and less than perfect information (Simon, 1955).

Decision-making theory also emphasizes the role of risk and uncertainty in choices. Digital twin technology has the potential to facilitate risk assessment and mitigation by providing a platform for scenario analysis and stress testing. Decision-makers can experiment with different scenarios and assess their impact on the physical system or process. This is consistent with the principles of prospect theory, which is based upon behavioral and economic theory that describes how people make decisions under uncertainty (Gremyr et al., 2019). Prospect theory has found applications in various decision-making processes, including public policy, healthcare, and organizational behavior (Camerer, 2005).

The relationship between decision-making theory and digital twins are intertwined and symbiotic. Decision-making theory provides the cognitive and theoretical framework for making effective choices, while digital twin offers the tools and technology to implement these choices in the real world. The integration of these two fields is evident in practical applications across various domains, enabling more informed decisions, improved risk management, and enhanced optimization.

The dynamic interplay between artificial intelligence (AI) and decision-making is perhaps one of the most significant domains in which AI is making a substantial impact. AI systems draw their roots from the advent of computers from which rule-based systems as expert systems evolved. Machine learning algorithms enable AI systems to learn from data, adapt and improve their performance over time. AI systems are designed to augment human decision-making processes by analyzing complex data sets and identifying patterns that may take humans longer to discern (Russell & Norvig, 2010). By integrating AI into decision-making processes, decision-making in complex and chaotic environments has the potential to provide more meaningful results. While AI systems should not be seen as replacements for the human, they should be seen as an integral part of the collaborative process of decision-making (Mittelstadt et al., 2016). Similarly, the digital twin does not replace the “real world” environment but provides a replica in which decisions may be explored from a wide spectrum of views and alternatives, providing a more rich and meaningful set of choices from which the decision-maker can choose.

The success of defense acquisition programs hinges on effective decision-making by program managers who must navigate budget constraints, technical intricacies, and evolving geopolitical landscapes (Schwartz, 2014). Inefficient decision-making can result in delays, cost overruns, and suboptimal outcomes, emphasizing the critical role of advanced technologies in augmenting decision-making capabilities. Systems engineering is a core methodology in the Defense Acquisition process and is often considered the principal method by which defense programs are managed and controlled. While systems engineering attempts to anticipate performance issues early in the acquisition life cycle, leading to improved system performance (Hossain & Jaradat, 2018), current methods are not agile enough to be able to provide reliable forecasting. Historical records bear this out by virtue of the many programs that continue to fail to meet their performance objectives and are often over cost and behind schedule.

Systems engineering primarily focuses on the technical aspects of a project, emphasizing the design, development, and integration of complex systems (Blanchard & Fabrycky, 2011). While this approach is invaluable for ensuring the functionality and reliability of military systems, it tends to neglect broader program management considerations such as budget constraints, political changes, and evolving threat landscapes (DoD, 2017). As a forecasting method, systems engineering may not adequately account for these external factors, leading to inaccurate predictions and suboptimal decision-making in defense acquisition.

Additionally, a key assumption of systems engineering is the stability of project requirements throughout the development life cycle (Blanchard & Fabrycky, 2011). However, in defense acquisition, requirements are often subject to change due to evolving geopolitical situations, emerging threats, and shifts in national security priorities (Oakley, 2019). The rigid nature of systems engineering may struggle to adapt to changing requirements, rendering it less effective as a forecasting method in the unpredictable landscape of defense acquisition.

Current methods of attempting to predict defense acquisition program performance are grounded in the theory of management that suggests that program monitoring and control can be accomplished through a performance measurement baseline (PMB) that measures work accomplished over time. Performance-Based Acquisition (PBA) is a strategic approach that emphasizes the measurement and management of performance outcomes. PBA aligns contracts, incentives, and milestones with desired performance levels. By using PBA, the DoD attempts to predict and influence performance throughout the acquisition process, ensuring that contracts incentivize suppliers to meet or exceed performance expectations. Techniques such as earned value management (EVM) draw heavily from performance measurement theories, particularly those related to assessing project progress and success. The concept of “earned value” itself is rooted in performance measurement, where the value of work performed is compared to the planned value to gauge project efficiency and effectiveness (Fleming & Koppelman, 2016). Additionally, Management Control Systems (MCS) theory, as developed by Robert N. Anthony (1965), emphasizes the need for organizations to implement systems that help manage and control their activities. EVM serves as a management control tool by providing a structured framework for measuring project performance against baselines, enabling proactive decision-making to address deviations from the plan (Fleming & Koppelman, 2016).

EVM is a performance measurement methodology that integrates scope, cost, and schedule to assess performance. It provides a structured approach designed to predict and control performance by comparing planned performance against actual progress. EVM is intended to offer insights into whether a program is on track to meet its performance goals (Fleming & Koppelman, 2016). EVM, however, does not account for changes in the overall development environment. It focuses on expected work relative to actual work and assumes that the risk and planning data adequately reflect the realities of the program. While this is a critical component of understanding a program, it is based upon preplanned cost and schedule information and does not provide sufficient clarity in prediction to be considered a viable forecasting method. It simply reaffirms planning processes and has little ability to anticipate the changing nature of the development environment.

New methods are required to better understand program performance and to be able to more reliably forecast program outcomes. This research will leverage theories and methods from other disciplines that provide a more accurate means of prediction. For example, digital twin theory is an emerging theory that is being used in disciplines to predict performance of discrete processes (Glaessgen & Stargel, 2012). Digital twin theory may provide new insight into the defense acquisition management process if coupled with complementary theories such as socio-technical or cyber physical theory, value theory and decision support theory. The concept of creating a virtual replica of a dynamic process such as defense acquisition presents the possibility of discovering the root causes of program challenges. Simply creating a replica of a process, however, is not sufficient. Insight and understanding will be gained through the intersection interpretation and alignment of complementary theories. Additionally, coherence in the model will be increased through the integration of artificial intelligence (AI) algorithms built into what will be referred to as the Acquisition Digital Environment (ADE). The integration of AI will create a learning model that improves its prediction ability as more data flows through the ADE.

Two concepts have emerged as integral components of current industrial and engineering processes: cyber physical systems and digital twins. These concepts represent a paradigm shift in how complex systems are designed, monitored, and optimized. Understanding these systems and the interrelationship between cyber physical systems and digital twins is an important step in being able to create an ADE that accurately represents the true nature of a human-centric business process. A cyber physical system refers to the integration of digital and physical elements where real-world entities interact with digital systems. Cyber physical systems have applications across many domains such as real-time monitoring and control and decision-making systems (Rathore et al., 2018). Digital twins are digital replicas of the physical world or systems. The virtual representation is informed by real-time data through sensors or other sources that provide a representation of the physical world. Digital twins can range from simple replicas to complex models of systems such as business and manufacturing processes to biological systems such as the human body (Glaessgen & Stargel, 2012).

Cyber physical and digital twin theory are related and mutually supporting in that cyber physical systems rely in real-time data for decision-making and control, which is also critical for the operation of digital twins. For example, smart manufacturing uses sensors on machinery to collect data that in turn is integrated into a digital twin to support decision-making and predictive maintenance to improve overall system performance (Tao et al., 2018). When applying this concept to a system such as the defense acquisition process, the cyber physical system is represented by the data collected through testing as well as cost and performance data of a program of record at various points during the developmental life cycle. Specifically, work breakdown structure data is the smallest unit of data of a program by which a program is measured. As a program moves through the life cycle, units of work are accomplished and can be measured and reported on through various program documents. In essence, this reporting data represent the sensor data that can be integrated into the digital twin of the life-cycle process. Digital twins facilitate simulations for understanding how physical systems operate under varying conditions. The data generated by cyber physical system is used to enhance these simulations, making them more accurate and reliable. The interconnectedness allows for the optimization of physical systems and processes in real or near real time (Tao et al., 2018).

Additionally, near real-time optimization of a physical system through a cyber physical system such as the ADE, may improve decision-making and forecasting. Real-time data-driven insight that is informed through learning systems such as artificial intelligence models further reveals the interconnectedness between cyber physical systems and digital twins (Rathore et al., 2018). Better forecasting of the defense acquisition process will improve decision-making, leading to better overall program performance. By forecasting program challenges before they become critical, decision-makers can react sooner and make more fiscally and programmatically sound decisions. Using cyber physical system such as an ADE, decision-makers can also "rehearse" scenarios before investing significant resources to improve acquisition strategy and planning.

Economic Value Theory

Decision-making is a fundamental aspect of human behavior that permeates all aspects of life, and its significance is particularly pronounced in the economic sphere. Current methods by which decisions in defense acquisition are informed are grounded in a cost-based approach that presumes to be able to forecast program performance relative to past estimates. This approach to decision-making excludes the concept of economic value as a key aspect to program performance and outcomes. Work, for example, is simply viewed as a function of cost and time rather than a unit of increasing value. By viewing work from a cost perspective rather than a value perspective, it is not possible to understand the underlying nature of why programs behave as they do. Decisions are focused on strict adherence to a cost-informed baseline, and the question of relevance in the form of value is often dismissed until programs are irreversibly dysfunctional, requiring new planning efforts to create new baselines. This process is inherently inefficient and leads to loss in operational capability to the customer and mismanagement of public resources.

Economic value theory provides a lens through which we can analyze the decisions individuals, businesses, and policymakers make, as these decisions ultimately shape the allocation of resources and contribute to the creation of value within an economy. Economic value theory is deeply rooted in classical and neoclassical economic thought. Adam Smith, often regarded as the father of modern economics, emphasized the role of individual decision-making in his seminal work, *The Wealth of Nations*. The concept of the invisible hand, where individuals pursuing their self-interest unintentionally contribute to the overall well-being of society, underscores the fundamental link between decision-making and economic value creation.

In the realm of microeconomics, consumer decision-making is key to economic value creation. Consumers make choices based on utility maximization, weighing the benefits and costs associated with different options (Samuelson, 1938). In the defense acquisition and program management context, value creation can be derived from work accomplished. If one decouples work from cost for the purpose of better definitizing value as something being created, each unit of work defined in the program work breakdown schedule then takes on a whole new utility by helping to change the framework of program choices from cost to value creation. The concept of marginal utility, derived from the law of diminishing returns, will play a pivotal role in explaining how the individual allocates resources to maximize satisfaction, therefore making decisions that are focused on maximizing value, not minimizing cost.

Value theory and decision-making are intrinsically linked and influence the choices humans make. The relationship between value theory and decision-making underscores the importance of understanding how individuals assess and prioritize values from both an economics and ethical perspective. Economic value theory provides the conceptual framework for assessing the worth of goods and services, while decision-making involves the process of selecting among competing alternatives. Classical economists such as Adam Smith emphasized the subjective nature of value while future economists included concepts such as individual preference and marginal utility (Mises, 1949). Similarly, decision-making involves assessing and prioritizing alternatives based on individual or collective preferences. The concept of utility, derived from economic value theory, plays an important role in decision-making in that individuals make choices from the perspective of perceived value of a set of options (Samuelson, 1938).

Additionally, economic value theory emphasizes the concept of opportunity cost where the value of the next best alternative is weighed against competing alternatives. Decision-makers weigh opportunity costs of their choices, considering what they must sacrifice in terms of alternative uses of resources. This perspective informs rational decision-making by encouraging individuals to select alternatives that maximize the overall best outcomes (Mises, 1949). This is precisely what program managers are supposed to be doing when making decisions; however, current methods do not provide adequate insight and nuance into the activities of a program under development to allow decisions to be optimized to be consistent with economic value and decision-making theory.

In physics, work is the energy transferred to or from an object via the application of force along a displacement. A force is said to do *positive work* if when applied in the direction of the displacement of the point of application. A force does *negative work* if it has a component opposite to the direction of the displacement (NCERT, 2020). For our purposes, positive work can be interpreted as an increase in overall value. The analogy between physical work and economic value will be developed as part of the overall research effort. However, for now we will draw upon the relationship that a unit of work identified in the program work breakdown structure can be used as a surrogate for a unit of value in that it represents a definitive action required to be accomplished at a specific time and that increasing positive work represents increasing positive value. As work is accomplished, this increases the economic value and provides an opportunity for program managers to weigh the value of the work performed at a certain point relative to the economic need at that moment in time. A virtual replica can simulate the flow of work increasing the choices available to the decision-maker. Additionally, decision tools will be able to learn as workflow is adjusted within the virtual replica, allowing for a smarter decision support system and over time increasing the potential for more accurate forecasting.

Socio-Technical Decision Support Systems

Decision and digital twin theory are becoming more intertwined through advances in technology, data analytics, and artificial intelligence. The two domains share a deep relationship in that decision-making theory focuses on understanding how individuals and organizations make choices, while digital twin theory centers on the creation of virtual replicas of processes that inform the decision-making process. Decision-making encompasses a wide range of concepts, models, and frameworks designed to explain and predict how humans, organizations, and increasingly, AI systems make choices when faced with uncertainty, risk, and complexity (Fraser, 1984). Various models such as rational choice model, bounded rationality, prospect theory, and the multi-attribute utility theory provide insight into the complexity of decision-making. Digital twin theory replicates the decision-making environment in that it aims to create a representation of the physical system or process in which decision theory can be applied.

Because of the interconnected nature of decision theory and digital twin theory, we must understand the relationships between the human and the technical framework within which decisions are made. The socio-technical influence on the decision-making process needs to be explored to make sense of key patterns that emerge during the development and testing of an ADE. As such, we must explore the current literature associated with the socio-technical theory as it relates to the decision-making processes. Socio-technical theory is a multidisciplinary framework that examines the interactions between social and technical elements within complex systems. Socio-technical theory originated in the field of organizational studies and has been applied in a variety of contexts such as information systems, health care, and transportation. Socio-technical theory can also be applied in the defense acquisition process in that the social and technical aspects of the process are interconnected and are considered a unified system which is influenced by context, culture, and the mutual influence of technology and human behavior (Trist & Bamforth, 1951). Trist and Bamforth argue that increasing technological advances integrated into the human centric process has shown to lead to improved efficiency and performance. We will explore whether increasing socio-technical integration through an AI-informed replica of the defense acquisition process has the same impact with regard to management decision-making and improved performance. Historical trends seem to suggest that there may be a positive correlation between the two. Many of the decision points within the defense acquisition process are informed by technologically driven data and require the integration of humans and technology to produce data that allows the process to function. Understanding this relationship is necessary to create an AI-informed digital replica of the defense acquisition process.

Additionally, socio-technical theory emphasizes the importance of human factors in technology and system design. A digital twin provides a platform for modeling and simulating human interactions with technology and data-driven processes that impact the overall system performance. This human-centric approach is necessary for ensuring that the technical solution or process design aligns with human needs, behaviors, and organizations priorities. To understand the significance of this, it is essential to explore how human-centric design fosters a harmonious relationship between technology and its users, thereby ensuring optimal functionality and long-term success. Adopting a user-centered approach enhances usability and overall satisfaction, ultimately contributing to the success of technological solutions (van Velsen et al., 2022). Human-centric design also extends to the organizational level. By aligning organizational priorities with design strategies, levels of efficiency can be gained by reducing waste and tasks that distract from value optimization. A study by Davenport (1993) in "Process Innovation: Reengineering Work Through Information Technology" emphasizes the need for technology to be an enabler of organizational objectives, emphasizing that a disconnect between technology and organizational goals can lead to inefficiencies. Davenport suggests that the business environment demands significant change and simply formulating strategy is no longer sufficient. He argues that we must also design the processes to implement it effectively by fusing information technology and human resource management to improve business performance (Davenport, 1993).

Understanding the socio-technical process involves recognizing the technological elements, such as hardware, software, and infrastructure, as well as the social aspects, including human participants, organizational structures, and culture. By comprehensively understanding these components and their interdependencies, one can provide a foundational framework for explaining the process (Trist & Bamforth, 1951). Socio-technical processes are dynamic and subject to continuous feedback and adaptation. Like other socio-technical processes, the defense acquisition process responds to feedback and changes and is intended to address issues, optimize performance, and align processes with evolving goals and objectives (Wong, 2022).

By replicating the defense acquisition process using a digital twin approach, greater clarity and alignment between organizational goals and outcomes can be achieved. The relationship between socio-technical and decision theory is inextricably linked and necessary to provide an opportunity to create a new way of examining defense acquisition that provides improved opportunities to forecast outcomes. Current management theory does not appear to reflect any research that attempts to replicate the defense acquisition process during the development phase of a program life cycle. This leaves us with a significant gap in knowledge on the relationship between the human and technical framework of the defense acquisition process that might provide insight into why defense programs under development tend to veer off track regarding meeting their cost, schedule, and performance objectives.

Technology Trust

The integration of Digital Twin technology and AI has revolutionized various industries, offering unprecedented opportunities for innovation and efficiency. However, the widespread adoption of these technologies is contingent upon the establishment of trust, both in the theoretical underpinnings and practical implementations. Equally important is gaining cultural acceptance of the technology to rely upon it to support critical decision-making. Due to the complexity and significant number of stakeholders in the defense acquisition process, the program manager is often influenced by competing information and priorities. Having a predictive model needs to account for the many priorities and stakeholder influences with a high degree of reliability for the capability to be adopted as a viable source by which to make decisions. Current decision support systems are slow and outdated due to the legacy data analysis approaches, such as EVM, that support these systems. An intelligent digital twin that allows for iterative scenario analysis and real-time forecasting would significantly improve the professional and theoretical foundation of program management.

Trust in digital twins and AI technology is anchored in a multidimensional theoretical framework that encompasses aspects of reliability, transparency, accountability, and ethical considerations. The theoretical foundation of trust in digital twins and AI hinges on their ability to provide reliable and accurate representations of the physical world. Digital twins, as virtual replicas of physical systems, must mirror reality while AI algorithms embedded in these twins must be capable of generating accurate predictions or simulations. These technologies are being used in other disciplines today with increasing success; however, they have never been used to replicate the program management process to improve program development performance outcomes. Product development leaders expect digital twins to accelerate product development processes and improve cost and schedule outcomes. Organizations are investing in the concept, with the global market for digital-twin technologies forecasted to grow at about 60% annually over the next 5 years, reaching $73.5 billion by 2027 (Argolini, 2023). This investment, however, does not represent a digitizing of the actual program management life cycle. It is predominantly focused on mechanical, test and evaluation, and manufacturing process to optimize these subordinate processes within the life cycle of a program under development (Argolini, 2023).

Trust in AI and digital twins requires transparency and explainability in their operations. Without good models and the right tools to interpret them, data scientists risk making decisions based on hidden biases, spurious correlations, and false generalizations. This has led to a rallying cry for model interpretability (Hohman et al., 2019). Theoretical frameworks such as interpretability in machine learning (Caruana et al., 2015) and explainable AI (XAI) methodologies emphasize the importance of making AI decisions understandable to humans. Transparency engenders trust by demystifying the nature of AI algorithms, fostering a clearer understanding of how decisions are made within digital twins. Creating a greater understanding of the intricate pathways with which information flows through the defense acquisition process, creating or reducing value, and gaining a sense of trust in the relationship between the digital version of reality will not only increase adoption of digital twin replicas of the defense acquisition process, but will further the theory and knowledge base of digital trust in general.

Scope

The defense acquisition process is a complex and multifaceted system designed to ensure the DoD acquires and sustains the most capable and cost-effective military capabilities. This process spans several distinct phases, each playing a critical role in the development, procurement, and sustainment of defense systems. Figure 1 shows a visual framework of the entire defense acquisition life cycle.

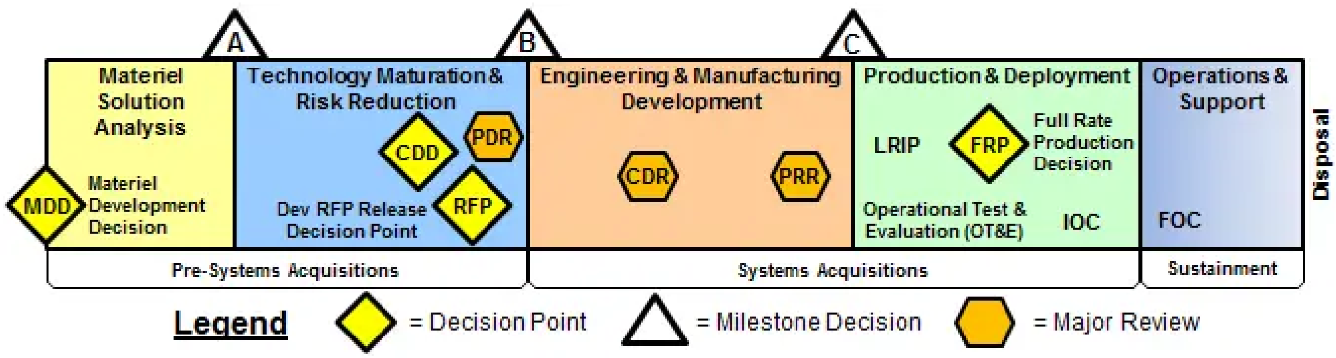


Figure 1. Defense Acquisition Life Cycle (AcqNotes)

The acquisition process begins with the identification of a capability gap or a need within the military, commonly referred to as a requirement. During the Concept and Technology Development phase, potential solutions are explored, and the feasibility of developing new technologies or adapting existing ones is assessed. This phase involves research, prototyping, and concept validation. Once viable concepts emerge, the Materiel Solution Analysis (MSA) phase focuses on refining and evaluating these options. This involves a detailed analysis of the potential solutions, cost estimates, and the development of an initial acquisition strategy. The goal is to identify the most suitable solution that meets the requirement within budget and time constraints. The Technology Maturation Risk Reduction (TMRR) phase is dedicated to advancing technologies and mitigating risks associated with the potential technology strategy. Prototypes are further developed and tested, and risk reduction measures are implemented. This phase aims to enhance the maturity of the technology and reduce uncertainties before entering development. The Engineering Manufacturing and Development (EMD) phase solidifies the design and development strategy and executes development of the product according to a budget and schedule baseline. Detailed engineering, manufacturing, and testing activities occur during this phase. The focus is on refining the design, ensuring manufacturability, and preparing for production. Successful completion of the EMD phase results in a system that is ready for production and deployment. This phase involves the mass production and deployment of the defense system. Manufacturing processes are optimized for large-scale production, and the system is fielded to military units. The Production and Deployment phase also includes ongoing logistics support to ensure the operational readiness and sustainment of the deployed systems (DoDI 5000.02, 2020)

This research will focus on the EMD phase of the defense acquisition process. While this phase might seem unique to the defense department, it is generalizable to the program management discipline at large. Choosing this phase is critical in that is represents the phase in which the organization commits to procurement of a capability and invests resources to its development. This phase also represents the phase in which program execution is grounded in prior predictions and measured against historical cost and schedule targets. In effect, this is the phase in which the "hype" for a project begins to decline either through poor program performance or ill-conceived requirements that were not well understood at the beginning of a project. This research will be divided into three phases. Each phase will culminate in a seminal paper that builds upon each other and supports the underlying research question.

1. Model Development – an initial digital twin model, which we will here forward refer to as the Program Management Acquisition Digital Twin (PMAD), will be developed that replicates the EMD phase of the life cycle. This model will be an exact replica with machine learning integrated into the model. A subset of artificial intelligence, the machine learning algorithms will learn from the data and improve performance automatically over time. This learning will be compared with past programs during phase two of the research to begin to establish a sense of confidence through increased reliability. This first paper will describe the process, development, validation, and complexity of building a new model of a complex business process system.
2. PMAD Forecasting Assessment – data from past programs will be used to mature the PMAD. Additionally, this phase will examine the actual information pathways relative to those predicted at program initiation and attempt to reveal root cause for differences. This root cause analysis will provide insight into relevant theory and practice that will extend the knowledge for these disciplines. Root cause understanding will also allow the machine learning tools to become more accurate in their predictions improving the potential for adoption of this process as a more accurate means of decision-making and forecasting. Additionally, data from the model will be compared to current decision-making and forecasting tools to assess the level of improvement in predicting program outcomes.
3. Current program assessment – the third phase of this research will take data from an existing program of record and use the PMAD to assess the approved acquisition strategy and program documents to provide insight into how the PMAD can improve real-world program performance and forecasting. This phase is necessary to increase the PMAD and underlying theory adoption. Additionally, this is a critical step in demonstrating generalizability of the results of this research across a larger population of program management environments.

Method

The defense acquisition process is a complex and resource-intensive endeavor that plays a pivotal role in ensuring the national security of the country. As technology advances, the integration of digital twin technology offers significant potential to enhance decision-making, efficiency, and cost effectiveness within the process.

This dissertation is a quantitative approach and will use an AI model developed for the purpose of this research to better understand the unstructured patterns of decision-making that occur within the defense acquisition process. The AI model will assess the patterns and decisions of actual historical programs of being replicated in a digital twin of the defense acquisition process. A digital twin is a virtual replica of a physical object or system that can be used for simulation, analysis, and testing. In the context of defense acquisition, an experiential twin can be used to simulate, analyze, and test real-world processes and experiences of the acquisition process to identify areas for improvement and optimize the process. The AI model helps to identify patterns within the defense acquisition process while simulating a program execution from the start of a contract award to a production decision. By integrating AI into a virtual twin of the defense acquisition framework, novel insight can be gained into how decisions are being made and the impacts of those decisions, leading to more informed decisions leading to improved outcomes of a Defense Acquisition Program of Record.

Research Design

- ***Phase I. Conceptual Framework and Model Development***

The defense acquisition process conceptual framework refers to the structured set of principles, guidelines, and procedures that govern how goods and services are acquired and managed for defense systems and capabilities. This process is crucial for ensuring that military organizations obtain the necessary resources and technologies to meet national security requirements effectively. The defense acquisition process typically involves several stages, and the conceptual framework provides a roadmap for decision-making and management throughout these stages.

Requirements Definition**:** The process begins with the identification and definition of the military capabilities needed to fulfill national security objectives. This involves assessing threats, analyzing operational needs, and specifying the desired capabilities.

Acquisition Planning**:** This stage involves developing a comprehensive plan for acquiring the necessary defense capabilities. It includes considerations such as budgeting, scheduling, risk management, and determining the acquisition strategy.

Contracting and Procurement**:** The acquisition process typically involves contracting with private companies or government agencies to design, develop, and produce the required defense systems. Procurement activities may include competitive bidding, negotiation, and contract award.

Development and Testing**:** Defense systems undergo design, development, and testing to ensure they meet performance, reliability, and safety standards. This stage may involve prototypes, testing in simulated environments, and eventually field testing.

Production and Deployment**:** Once a defense system successfully completes testing, it enters the production phase. This stage involves manufacturing the systems in larger quantities and deploying them to military units.

Life-Cycle Management**:** The defense acquisition process extends beyond initial deployment. It includes ongoing maintenance, upgrades, and eventual retirement or replacement of systems as they become obsolete or reach the end of their operational life.

Regulatory Compliance**:** Throughout the acquisition process, adherence to legal and regulatory frameworks is crucial. This may include compliance with procurement laws, export controls, and other relevant regulations.

Cost Management**:** Cost considerations play a significant role in defense acquisition. This includes estimating, budgeting, and managing costs throughout the life cycle of a defense system.

Risk Management**:** Identifying and managing risks is essential to ensure the success of the acquisition process. This involves assessing potential challenges and implementing strategies to mitigate or address them.

While the defense acquisition process is designed to be systematic, transparent, and accountable, its complexity lends itself to a process that is difficult to manage due to the lack of deep understanding of the interdependencies of the ties between the elements of the framework from both a human and technical perspective. A virtual replica model of the defense acquisition process will be constructed using relevant data sources, including historical acquisition data, stakeholder interviews, and existing process documentation. While the model itself is not the primary intent of this research, it is necessary to understand the complex socio-technical environment of the defense acquisition process. By developing a virtual replica that mimics the "real world," the researcher will be able to gain insight into how information flows through the process and how decisions are made. This in turn will help to improve the model and eventually facilitate the creation of a tool that can improve forecasting and provide a reliable method by which program performance can be predicted. Key steps in developing the model include:

Define Objectives and Scope - Clearly define the objectives of the virtual replica model. Determine the specific aspects of the defense acquisition process you want to simulate. Identify the scope of the model, including the stages of acquisition, key stakeholders, decision points, and relevant environmental factors.

Gather Data and Information - Collect data on the defense acquisition process, including historical information, regulations, policies, and procedures. Consult with subject matter experts (SMEs) who have experience in defense acquisition to ensure accuracy and relevance.

Identify Key Components and Relationships - Break down the defense acquisition process into its key components, such as requirements definition, contracting, testing, and deployment. Identify the relationships and dependencies between these components to accurately represent the flow of activities.

Select Modeling Tools and Develop Model - Choose appropriate modeling tools for creating the virtual replica. This could include simulation software, 3D modeling tools, data visualization platforms and artificial intelligence models. Define the logic and rules governing the interactions between different elements in the virtual environment.

The model will be designed to capture the dynamic and interconnected nature of the defense acquisition life cycle.

- ***Phase II. PMAD Forecasting Assessment***

The virtual replica model will be subjected to simulations using real-time and historical data. Validation of the model will be conducted by comparing the simulation results with actual acquisition outcomes. This step aims to ensure the accuracy and reliability of the digital twin in replicating the complexities of the defense acquisition process. Having a validated virtual replica that can be used to inform decision-making and resource priorities has many benefits such as:

Variables and Data Analysis

In program management, variables can significantly impact the effectiveness of program performance. These variables are often interconnected and need to be carefully managed to ensure successful outcomes. While variables such as quality, communication, and leadership are critical to program performance, the purpose of this research is to explore the ability to forecast outcomes. Program performance outcomes are typically measured by a program's cost and schedule during execution relative to the initial program plan. Currently, the principal method by which major development programs are measured in terms of program execution is EVM, a project management technique that integrates cost, schedule, and scope to assess project performance and progress. It provides a standardized and objective method for project managers to measure a project's performance against its baseline plan. EVM is typically used for tracking and forecasting project costs and schedule performance.

Rather than focusing on traditional metrics, this research will measure a program's performance through value and workflow. Every development program starts with a pre-approved work breakdown structure that represents work over time. Each of these work packages was developed according to a specific criterion designed to complete the project in a specified period of performance. If the planned work is correct and it is performed exactly according to plan, then the output meets the needs of the customer and can be said to provide 100% value. As a unit of work is performed, it provides a unit of value to the program. If this unit of value is performed at a predetermined time, this is the highest level of efficiency a program might be able to achieve. We will need to assess all the units of work simultaneously over time, comparing actual to planned to provide a more accurate assessment of how much value a program is generating. The key variables we will use for comparison to current methods are:

Dependent Variable: Value (expected work completed relative to planned outcomes throughout the program life cycle)

Independent Variable: Workflow (Work Breakdown Structure)

The reality is that no program accomplishes every work package according to the work breakdown structure plan, resulting in program inefficiencies and loss. Additionally, there is no method available to assess program value and subsequently generate accurate program performance predictions. By using units of work to inform value, we can decouple cost from the analysis and focus on value and time (Schedule) relative to workflow. A sample two by two comparison helps illustrate how different scenarios or approaches may be classified based on these dimensions.

	High Value	**Low Value**
Efficient Workflow		
Inefficient Workflow		

- High Value (Efficient Workflow) - Programs in this quadrant exhibit both high value and efficiency in workflow. They deliver significant value within a relatively short timeframe, indicating effectiveness in resource utilization and project management.
- Low Value (Efficient Workflow) - This quadrant represents programs that deliver high value but may take a longer time to complete. While the workflow is efficient, the complexity or scope of the program requires a more extended timeline for successful execution.
- High Value (Inefficient Workflow) - Programs in this quadrant have a quick workflow but deliver low value. The focus may be on completing tasks rapidly, but the overall impact or benefits are limited.
- Low Value (Inefficient Workflow) - This quadrant represents programs with both low value and an inefficient workflow. These programs may take a long time to complete, and the outcomes may not justify the time and resources invested.

The virtual replica/AI model developed for this research will allow for an analysis of large value variables simultaneously and will provide the structure by which we can learn from actual data, creating the conditions for a predictive tool that will enhance program management insight and decision-making.

Validating a digital model of a program management process often involves the use of statistical tools to assess the accuracy, reliability, and effectiveness of the model. Validation will begin by validating the requirements for the model. Ensuring that the twin accurately represents the key elements of the program management process and aligns with the goals set by stakeholders is critical to ensuring the model is an accurate representation of the actual process. Additionally, testing of the digital twin under various scenarios and conditions and comparing the digital twin's predictions with actual outcomes or historical data will help to verify its reliability.

Using Value as the dependent variable, a comparison will be made to EVM performance metrics to assess correlation between the model and actual program performance. This will help assess how well the twin replicates the behavior and outcomes of real-world scenarios, providing a basis for validating its reliability. Statistical tools such as regression analysis, hypothesis testing, and goodness-of-fit tests will help to evaluate the performance of the digital twin and assess whether the twin's predictions align with the observed data of actual program performance. Additionally, a sensitivity analysis to understand how changes in work package input parameters impact the outputs of the digital twin will identify the robustness of the twin under varying conditions, contributing to its reliability.

This research is focused on understanding knowledge from new methods that will improve the decision support process of program management. Therefore, once the validity of the digital twin is sufficiently determined, the data from the model will be used to compare with existing methods of decision-making and forecasting in development programs. Current methods used to manage development programs are based upon EVM techniques, which produce key performance variables that are based on cost, schedule, and work performed. EVM is a project management technique that integrates cost, schedule, and scope to assess project performance and progress. It provides a standardized and objective method for project managers to measure a project's performance against its baseline plan. EVM is typically used for tracking and forecasting project costs and schedule performance.

Data

Data for the digital twin model will be sourced from a variety of channels, including historical defense acquisition data from relevant government agencies, stakeholder interviews, and expert opinions to capture qualitative insights, and technical specifications and documentation related to defense acquisitions. Historical defense acquisition data will primarily be drawn from existing program work breakdown structure data, in that this represents the lowest level of work defined for a program under development. This data represents a sample of the type of data available and the level of detail that will inform the research process. Other data sets will include program schedule data, earned value metrics, and selected acquisition reports.

Ethical Considerations

This research will adhere to ethical guidelines, ensuring the confidentiality of sensitive information and obtaining informed consent from stakeholders involved in interviews. All data will be anonymized to protect the identity of individuals and organizations. An IRB determination will be obtained prior to conducting any interviews.

Expected Results

The anticipated results include a validated virtual replica of the defense acquisition process that provides improved decision-making methods and a level of forecasting that does not currently exist in the program management profession. Additionally, this research will provide insight into potential improvements of the defense acquisition process facilitated by a virtual replica, and recommendations for implementing this technology in defense acquisition strategies. This dissertation's methodology combines literature review, conceptual framework development, data collection, simulation, and analysis to explore the application of digital twins in studying and improving the defense acquisition process. The results of this research aim to contribute to a more efficient, transparent, and adaptive defense acquisition framework.

This study is important because there is currently no useful simulation environment in the defense acquisition process that allows program managers to virtually execute their programs and observe them under conditions of ambiguity and change. By allowing the PM to simulate program volatility before a program has even begun, decisions can be made prior to program disruption, improving the probability of program success.

References

Anthony, R. N. (1965). *Planning and control systems: A framework for analysis*. Harvard University Press.

Argolini, F. (2023). *Digital-twin technologies can help companies create products faster. They could transform the work product development too*. McKinsey & Company https://www.mckinsey.com/industries/industrials-and-electronics/our-insights/digital-twins-the-key-to-smart-product-development

Argolini, R. (2023). *Digital twin: The key to smart product development*. McKinsey Global Publishing, McKinsey & Company.

Blanchard, B. S., & Fabrycky, W. J. (2011). *Systems engineering and analysis*. Pearson.

Camerer, C. F. (2005). Three cheers—psychological, theoretical, empirical—for loss aversion. *Journal of Marketing Research, 42*(2), 129–133.

Choo, C. W., & Bontis, N. (Eds.). (2002). *The strategic management of intellectual capital and organizational knowledge*. Oxford University Press.

Davenport. (1993). In *Process innovation: Reengineering work through information technology*. Harvard Business Press.

Department of Defense Instruction. (2020). *Operation of the adaptive acquisition framework*.

DoD. (2017). *Defense acquisition guidebook*.

Elkins, M. (2019). *Defense acquisition: Observations on the National Defense Authorization Act for Fiscal Year 2016*. GAO.

Fleming, Q. W., & Koppelman, J. M. (2016). *Earned value project management*. Project Management Institute.

Fox, R. (2016). *Defense acquisition reform, 1960–2009: An elusive goal*. Stanford University Press.

Fraser, R. B. (1984). Decision analysis and artificial intelligence. *Artificial Intelligence*, *24*(2), 263–297.

Glaessgen, E. H., & Stargel, D. S. (2012). The digital twin paradigm for future NASA and U.S. Air Force vehicles. *Proceedings of the 53rd AIAA Structural Dynamics and Materials Conference*.

Hohman, F., et al. (2019). *Gamut: A design probe to understand how data scientists understand machine learning models*. CHI 2019, May 4–9, 2019, Glasgow, Scotland.

Hossain, N., & Jaradat, R. (2018). *Proceedings of the American Society for Engineering Management 2018 International Annual Conference E-H*. Ng, B. Nepal, E. Schott, and H. Keathley, eds.

IBM. (2023). *What is a digital twin?* https://www.ibm.com/topics/what-is-a-digital-twin#:~:text=A%20digital%20twin%20is%20a,reasoning%20to%20help%20decision%20making

Mises, L. (1949). *Human action: A treatise on economics*. Yale University Press.

NCERT. (2020). Physics Text, Ch 6.

Oakley, S. (2019). *Defense acquisition: Observations on the National Defense Authorization Act for Fiscal Year 2016*. GAO.

Project Management Institute. (2017). *A guide to the Project Management Body of Knowledge (PMBOK Guide)*.

Rathore, H., Uber, T., & Shah, M. A. (2018). Cyber-physical systems and digital twin: Comprehensive review. *IEEE International Conference on Intelligent Robots and Systems (IROS)* (pp. 3719–3726).

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, *5*(17), 61–71.

Schwartz, M. (2014). *Defense acquisition: How DOD acquires weapon systems and recent efforts to reform the process*. Congressional Research Service.

Scott, J., et al. (2014). A decision support system for supplier selection and order allocation in stochastic, multi-stakeholder and multi-criteria environments. *International Journal of Production Economics.*

Simon, H., & McSweeney, L. (2010). A behavioral model of rational choice. *Competition Policy International*, *6*(1).

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*(1), 99–118

Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations*. W. Strahan and T. Cadell.

Sullivan. (2017). *Defense acquisitions: Assessments of selected weapon programs*. GAO.

Sullivan, M. (2009). *Measuring the value of DOD's weapon programs requires starting with realistic baselines* (GAO-09-543T). GAO.

Tao, F., et al. (2018). Digital twin in industry: State of the art. *IEEE Transactions on Industrial Informatics*, *15*(4), 2405–2015.

Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the Longwall method of coal-getting: An examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. *Human Relations*, *4*(1), 3–38. https://doi.org/10.1177/001872675100400101

van Velsen, L., Ludden, G., & Grünloh, C. (2022). The limitations of user- and human-centered design in an ehealth context and how to move beyond them. *Journal of Medical Internet Research*, *24*(10), e37341.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.

Weyn, J. A., Durran, D. R., & Caruana, R. (2019). Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *Journal of Advances in Modeling Earth Systems*, *11*, 2680–2693. https://doi.org/10.1029/

Wong, P., et al. (2022). *Improving defense acquisition*. RAND.

Introducing SysEngBench: A Novel Benchmark for Assessing Large Language Models in Systems Engineering

Ryan Bell—is an 8-year experienced engineer in the defense industry. In his current role at Naval Information Warfare Center Atlantic (NIWC LANT), Bell provides modeling and simulation expertise to a variety of programs for the Navy and USMC. He specializes in simulating communication systems in complex environments and is an advocate for the use of digital engineering early in the systems engineering lifecycle. Bell earned a BS in electrical engineering from Clemson University and an MS in electrical engineering from Clemson University with a focus on electronics; he is currently pursuing his PhD in systems engineering at the Naval Postgraduate School. He is a South Carolina registered Professional Engineer (PE), published author, and teacher.

Ryan Longshore—is an 18-year veteran of both the defense and electric utility industries. In his current role at Naval Information Warfare Center Atlantic (NIWC LANT), Longshore leads a diverse team of engineers and scientists developing and integrating new technologies into command and operations centers. Longshore is heavily involved in the Navy's digital engineering transformation and leads multiple efforts in the model based systems engineering and model based engineering realms. Longshore earned a BS in electrical engineering from Clemson University and an MS in systems engineering from Southern Methodist University; he is currently pursuing his PhD in systems engineering from the Naval Postgraduate School. He is a South Carolina registered Professional Engineer (PE), an INCOSE Certified Systems Engineering Professional (CSEP), and has achieved the OMG SysML Model Builder Fundamental Certification.

Raymond Madachy, PhD—is a Professor in the Systems Engineering Department at the Naval Postgraduate School. His research interests include system and software cost modeling, affordability and tradespace analysis, modeling and simulation of systems and software engineering processes, integrating systems engineering and software engineering disciplines, and systems engineering tool environments. His research has been funded by diverse agencies across the DoD, National Security Agency, NASA, and several companies. He has developed widely used tools for systems and software cost estimation and is leading development of the open-source Systems Engineering Library (se-lib). He received the USC Center for Systems and Software Engineering Lifetime Achievement Award for "Innovative Development of a Wide Variety of Cost, Schedule and Quality Models and Simulations" in 2016. His books include Software Process Dynamics and What Every Engineer Should Know about Modeling and Simulation; he is the co-author of Software Cost Estimation with COCOMO II and Software Cost Estimation Metrics Manual for Defense Systems. He is writing Systems Engineering Principles for Software Engineers and What Every Engineer Should Know about Python.

Abstract

In the rapidly evolving field of artificial intelligence (AI), Large Language Models (LLMs) have demonstrated unprecedented capabilities in understanding and generating natural language. However, their proficiency in specialized domains, particularly in the complex and interdisciplinary field of systems engineering, remains less explored. This paper introduces SysEngBench, a novel benchmark specifically designed to evaluate LLMs in the context of systems engineering concepts and applications. SysEngBench will encompass a comprehensive set of tasks derived from core systems engineering processes, including requirements analysis, system architecture design, risk management, and stakeholder communication. By leveraging a diverse array of real-world and synthetically generated scenarios, SysEngBench aims to provide an assessment of LLMs' ability to interpret complex engineering problems and generate innovative solutions.

Our evaluation of leading LLMs using SysEngBench reveals significant insights into their current capabilities and limitations in systems engineering contexts. The findings suggest pathways for future research and development aimed at enhancing LLMs' utility in the systems engineering discipline. SysEngBench contributes to the understanding of AI's potential impact on systems engineering.

Keywords: Systems Engineering, Large Language Models (LLMs), Benchmark, SysEngBench, Performance Evaluation, Intelligent Decision Making

Introduction

The intersection of artificial intelligence (AI) and engineering presents a frontier with the potential to revolutionize how we approach complex engineering challenges. One field that focuses on architecting solutions for complex engineering challenges is systems engineering, an emerging engineering field that can capitalize on the widespread proliferation of AI to mature the field at a more rapid pace. In order to harness AI, an understanding must be established on how well Large Language Models (LLMs) perform within the field - an understanding that is not yet baselined for systems engineering. This paper seeks to target this knowledge gap with a targeted approach to assess the capabilities of LLMs within the domain of systems engineering.

This paper introduces SysEngBench, a pioneering benchmark designed to assess LLMs against a diverse set of concepts and applications encountered in systems engineering. The motivation behind SysEngBench stems from the recognition that there has been an evolution of benchmarks from common sense, to inference, to field specific. Evaluation of LLMs within field specific domains has already begun - from high school courses to medical exams to law exams (sources). As LLMs become more capable, more complex benchmarks must be made to continue to track progress. As benchmarks become more complex, field specific knowledge is necessary from practitioners and experts in the field. SysEngBench is the proposed benchmark for the systems engineering field and seeks to incorporate field practitioners and expert knowledge. The proposed framework is not all encompassing nor complete at this time of writing and seeks feedback from the community. More specifically, the objective of this paper is to provide the initial concept and framework of the benchmark to be molded.

Background and Related Work

Overview of Systems Engineering

Systems engineering stands at the convergence of various engineering disciplines, aimed at developing coherent and effective systems through a system lifecycle process. It involves methodologies and practices that ensure all aspects of a system's lifecycle, from conceptualization to decommissioning, are considered and optimized. This interdisciplinary approach addresses complexity by emphasizing robust planning, design, analysis, and management practices. Various methodologies are used within the systems engineering community, from the traditional systems engineering "vee" model, to the spiral model, to the waterfall model, and several others (Boehm, 1986).

Systems engineering spans across industries, including aerospace, automotive, software, and more, making systems engineering the glue to stitch together all of the other fields. In recent years, significant progress has been made with respect to Model Based Systems Engineering (MBSE) tools. These modeling and simulation tools help to understand the interlinking between industries and effectively manage the available trade space for any given system or system of systems modeled. MBSE tools sit at the intersection of modeling languages, structure, model based processes, and presentation frameworks.

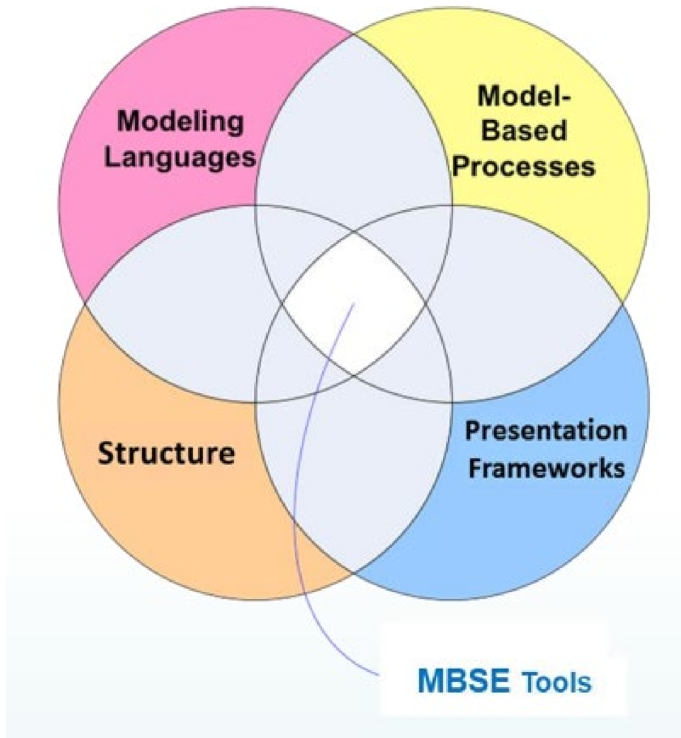


Figure 12. Model Based Systems Engineering Venn Diagram (Vaneman, 2020)

Front running tool sets include the likes of Cameo Enterprise Architect, Innoslate, among others. Most run on a UML backbone modified for systems engineering called SysML. Recent advancements have been made to SysML, known as SysMLv2, as an effort to democratize and open source systems engineering modeling. The architecture of SysMLv2 – which includes a textual format – will allow for a more fluid ability to train LLMs on models in the field.

The traditional systems engineering lifecycle is quite document-centric. In recent years, there has been a push to move towards model-centric management of the systems engineering lifecycle. Document-centric focuses on generating documents and those documents being the authoritative sources of truth for each of the milestones and associated efforts within the lifecycle, leading to an increasingly disaggregate pile of information – where sorting through this information to get a complete picture of how requirements and relationships within the system are represented also becomes increasingly complex. Due to the sheer amount of information and documentation, LLMs could significantly reduce the cognitive load and increase understanding of a systems current stature within the lifecycle, especially when aggregated into a single source of truth model (Defense Acquisition University [DAU], 2024; *SEBoK*, 2024).

Review of LLMs

Large Language Models (LLMs) such as GPT-4 have revolutionized the field of natural language processing (NLP) by demonstrating an impressive ability to understand and generate human-like text. These models are trained on vast amounts of text data, enabling them to grasp a wide range of topics, infer context, and produce coherent and contextually relevant responses. LLMs have been applied in numerous applications, from writing assistance and chatbots to more complex tasks like code generation and summarization.

Within the context of types of LLMs, there are different levels of accessibility, training sources, and varying levels of fidelity. For accessibility, there are open source models like Llama 2, Falcon, and Dolly as well as proprietary models like GPT-4, Claude, and Bard. Open source models are available in various repositories – one of the largest being HuggingFace. In general, proprietary models have been outperforming open source models, but the gap continues to close on the leaderboards. Every model is trained in a different set of data sources – some scrape GitHub for code, some scrape wikis and other openly available information or textbooks, and others are trained on private corpuses.

When it comes to fidelity, there are different preferences for fidelity based on the available hardware. A technique called quantization is common, where inference is ran on lower precision data types than the usual 32-bit (HuggingFace, 2024b). While this does result in lower fidelity, one can then run the model more easily on local hardware (Talamadupula, 2024). An example of a model released at varying data sizes is Llama 2, available in 7B, 13B, and 70B (*Llama 2: Open Foundation and Fine-Tuned Chat Models*, 2024).

To use models, different prompt can be used to change the output, a strategy described as prompt engineering. Prompts range from zero shot (no context provided and one try), to few shot (x number of refining attempts), and to Retrieval Augmented Generation (RAG). RAG pulls in relevant information from a large corpus of data at inference time when indexed properly. If prompt engineering does not give the desired answer, further fine-tuning of the model is required. Custom fine-tuning of LLMs on domain-specific datasets can significantly enhance their performance on specialized tasks.

Existing Benchmarks

The landscape of AI benchmarks has evolved over time, with early benchmarks focusing on foundational tasks such as work relationships and their semantic similarities to more recent, increasing complexity benchmarks such as College Medicine, Physics, Biology, Computer Science, Math, Electrical Engineering, among others (Hendrycks et al., 2021). Other non-technical outputs of LLMs are also being studied. The progression of increasing complexity is demonstrated in the table below, which shows the benchmark name, topic of the benchmark and the date the benchmark was initially released (*AI Fundamentals*, 2023). The list is not meant to be all encompassing or a review of literature, but rather a brief look at the evolution of benchmarks and their purpose over time.

Table 2. LLM Benchmarks over Time

Benchmark Name	Topic	Released	Type of Benchmark
WordNet	Word relationships and meanings, foundational dataset for semantic similarity and language understanding	1985	Natural Language Processing
MNIST	Handwritten digit recognition, foundational for image processing and computer vision	1998	Image Processing
BLEU	Language translation quality metric, foundational for evaluating machine translation systems	2002	Natural Language Processing
Enron Emails	Recognizing names, entities, and information extraction from natural email datasets	2004	Natural Language Processing
ImageNet	Large-scale image recognition and classification, pivotal in advancing deep learning in computer vision	2009	Image Processing

LAMBADA	Understanding context and reasoning in text, focusing on predicting sentence endings (Paperno et al., 2016)	2016	Natural Language Processing
SWAG	Common sense reasoning and predicting plausible sentence endings in a given context (Zellers et al., 2018)	2018	Natural Language Processing
GLUE	A collection of diverse NLU tasks like question answering and sentiment analysis to advance language understanding across various contexts.	2018	Natural Language Processing
SuperGLUE	A successor to GLUE with more challenging tasks, pushing the limits of NLU models with advanced reasoning and co-reference resolution.	2019	Natural Language Processing
HellaSWAG	An extension of SWAG for more challenging common sense reasoning scenarios (Zellers et al., 2019)	2019	Natural Language Processing
ARC	"ARC evaluates an AI's ability to tackle each task from scratch, using only the kind of prior knowledge about the world that humans naturally possess, known as core knowledge" (Clark et al., 2018; Lab42, 2024).	2019	Natural Language Processing
DROP	Reasoning over paragraphs, requires numerical reasoning and understanding of natural language (Dua et al., 2019)	2019	Natural Language Processing
Winogrande	A large-scale dataset of winograd schemas designed to improve commonsense reasoning in AI systems	2019	Natural Language Processing
XTREME	Cross-lingual understanding and translation across multiple languages, tests multilingual capabilities	2020	Natural Language Processing
MMLU	Measures professional and academic knowledge across various fields including College Medicine, Physics, Biology, Comp Sci, Math, Electrical Engineering, Professional Accounting, Psychology and worldly knowledge about Foreign Policy and Religions, among others (Hendrycks et al., 2021)	2021	Natural Language Processing
TruthfulQA	A question-answering dataset designed to evaluate a model's ability to produce truthful and factual answers.	2021	Natural Language Processing
GSM8K	Grade School Math 8K (GSM8K), a collection of math word problems aimed at evaluating numerical reasoning	2021	Natural Language Processing
BIG-Bench	Broad spectrum of tasks testing reasoning, common sense, professional knowledge, and language capabilities (*Google/BIG-Bench*, 2021/2024)	2022	Natural Language Processing

Performance of models across benchmarks are available in various locations, with the de-facto location being HuggingFace's (2024a) leaderboard. A list of other leaderboards is available on the site.

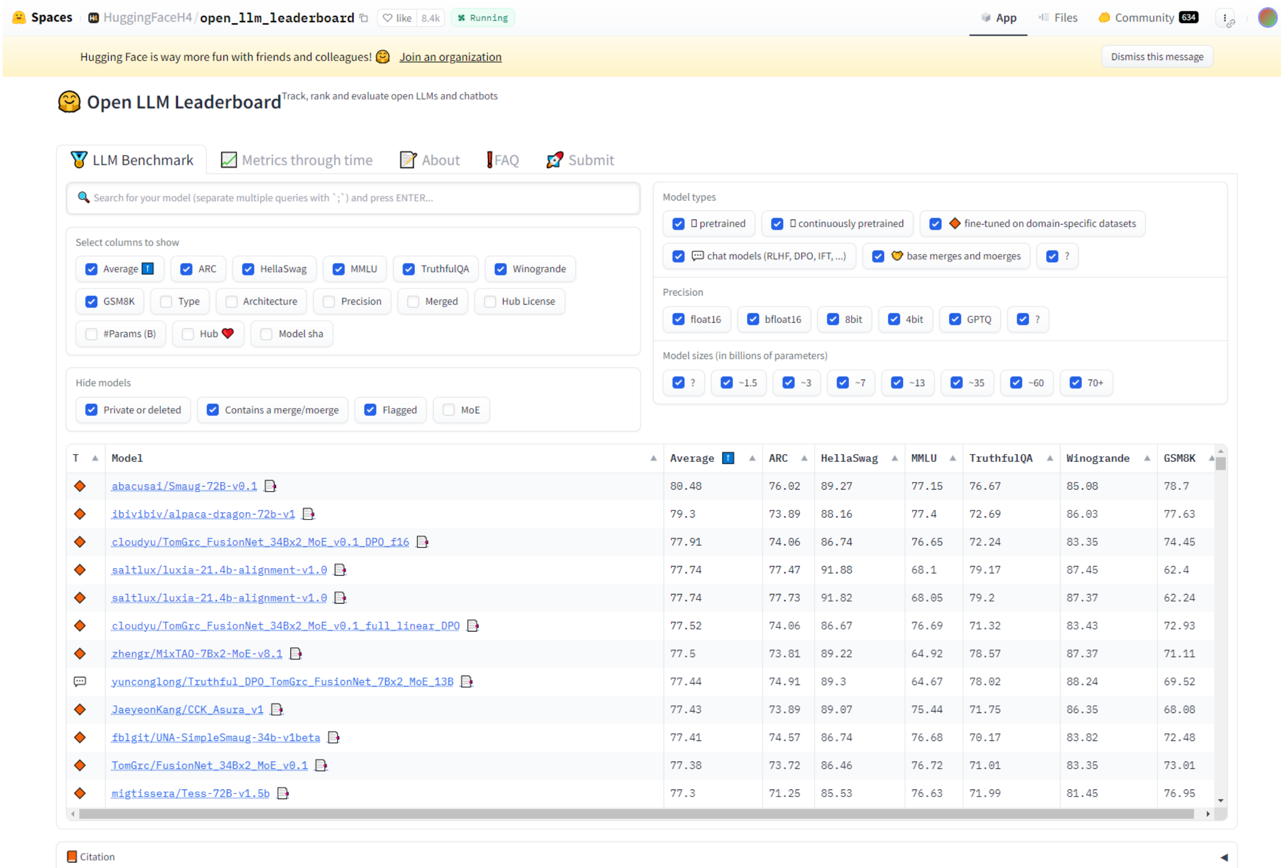


Figure 13. HuggingFace Leaderboards Screenshot

Benchmark Frameworks

Benchmarks range significantly when evaluating a domain specific field and level of complexity within that domain. Benchmarks can take a simple question and answer format, a multiple choice format, a fill in the blank format, an open ended response format, or various other methods. The more clear the answer has to be, the more clear the evaluation of a model with a given benchmark. Other more soft metrics are used for non-definitive answer scenarios to measure “correctness.” This can range from measuring token counts, biases, tone, or otherwise.

Data Sources and Generation

Data sources for language models can vary widely. Some include professional documents only from journal articles and textbooks, while others also ingest blog posts and other sources, but one thing that remains common across all language models is that garbage in equals garbage out.

With regards to benchmark generation, datasets can be completely synthetic, semi-synthetic, or done completely by hand. Perhaps the worst quality assurance (QA) process is full synthetic, although for various types of data, this may be within acceptance criterion and the best method for creating data at scale (Lambert, 2023; Packt, 2024; *Synthetic Data*, 2024). For a domain specific application, semi-synthetic or by hand is recommended for the highest fidelity.

SysEngBench: Framework and Design

SysEngBench Framework

The framework selected for SysEngBench is a simple multiple choice question benchmark. The benchmark currently covers an introduction to systems engineering but will be expanded to sub-fields within systems engineering discussed in future work. The current fundamentals of systems engineering questions are questions that should be correctly answered by graduate level systems engineering students at least 1 year into their course work at the Naval Postgraduate School.

SysEngBench Categories

SysEngBench selects 10 topics to reflect the core processes of systems engineering. The 10 main areas and their descriptions can be found in **Table 18** below.

Table 2. SysEngBench Topics

Area	Description
Requirements	Tasks that simulate the extraction, interpretation, and validation of system requirements from diverse sources, including stakeholder interviews and technical documents.
System Architecture and Design	Tasks that involve designing system architectures, considering aspects like modularity, scalability, resilience, and integration with existing systems.
Model-Based Systems Engineering (MBSE)	Tasks focusing on the application of modeling approaches to support system requirements, design, analysis, verification, and validation activities throughout the system lifecycle.
Cost Modeling	Tasks related to estimating, analyzing, and optimizing costs associated with the development and deployment of complex systems, taking into account factors such as materials, labor, and operational expenses.
System Capability/Suitability Engineering (-ilities)	Tasks aimed at evaluating and enhancing the overall performance and suitability of systems, including assessments of reliability, maintainability, and other critical 'ility' factors that affect system effectiveness and lifecycle cost.
Safety Engineering	Tasks involving the identification and mitigation of hazards, as well as the analysis of potential safety risks to minimize the likelihood and impact of accidents and failures.
Human Factors Engineering	Tasks that consider the interaction between humans and systems, aiming to optimize system performance via user interfaces, ergonomics, and usability studies.
System Integration and Development	Tasks focusing on the process of bringing together system components into a whole and ensuring that those components function together as intended, addressing challenges in integration and interoperability.
System Verification and Validation (V&V)	Tasks related to the confirmation that a system meets defined specifications and requirements (verification) and that it fulfills its intended purpose (validation), involving a combination of testing, analysis, and review techniques.
Risk Management	Tasks that require identifying potential risks, assessing their impact, and devising mitigation strategies, crucial for ensuring system reliability and safety.

The current iteration of the benchmark does not include all of the topic fields above since scope was currently limited to SE 3100 but is what will be strived for with future iterations, including refactoring the fundamentals tested into the proposed SysEngBench topics.

Table 3. Benchmark Question Distribution

Row Labels	Question Count	Question %
Fundamentals of SE	**116**	**100.00%**
SE Definitions	9	7.76%
Problem Definition and Stakeholders	11	9.48%
MBSE Overview	4	3.45%
Requirements	22	18.97%
Functional Analysis	11	9.48%
Value System Design	13	11.21%
Architecture	6	5.17%
Decision Making	10	8.62%
Risk	3	2.59%
System Integration, Qualification, Costs, Life Cycle Issues	27	23.28%
Grand Total	**116**	**100.00%**

SysEngBench Data Sources and Generation

The data sources used included lecture slides from SE 3100 at the Naval Postgraduate School. The syllabus for the class includes the following knowledge gained after taking the course:

- Define systems engineering, including its purpose and scope and the role of the systems engineer.
- Define systems architecting, including its purpose and scope and the role of the systems architect.
- Apply the fundamentals of a systems engineering process appropriately across a system's lifecycle.
- Elicit, elaborate and document system requirements based on user needs and operational objectives; translate them to technical requirements.
- Create a system value hierarchy reflective of stakeholder goals.
- Complete system functional analysis in support of requirements engineering using modeling tools such as IDEF0, FFBD and other techniques.
- Develop, evaluate and document alternative system architectures. A supplemental joint effort throughout the course will be to gain a common understanding of the applications of Systems Engineering in the Department of Defense (DoD).

The multiple choice questions were created with some AI assistance, but each was reviewed by a human systems engineer for correctness for a semi-synthetic dataset. More complex questions will investigate the LLMs ability to reason "within the gray" of systems engineering, particularly higher dimensional trade spaces where there are multiple configurations that would meet requirements.

Implementation and Benchmarking Process

Model Selection

A few common open source LLMs were selected for their availability and to show a range of performance. The models selected were Llama 2, Mistral, and Orca 2. All models used 8 bit quantization. The largest quantization available that would fit on a 32GB (or 64GB machine) was selected for each model.

Table 4. Models Used

Source	Model	Size	Quantization
TheBloke	Orca-2-7B-GGUF	7.16GB	8 bit
TheBloke	OpenHermes-2.5-Mistral-7B-GGUF	7.70GB	8 bit
TheBloke	Llama-2-7B-Chat-GGUF	7.16GB	8 bit

Benchmarking Procedure

The benchmarking process for SysEngBench is designed to be modular and replicable, as well as run locally or via cloud LLMs for future tests. To get some quick results, a simple evaluation method of querying for a response and parsing for a letter choice was implemented. To push for repeatability and scalability for future tests, lm-evaluation-harness will be implemented (*EleutherAI/Lm-Evaluation-Harness*, 2020/2024).

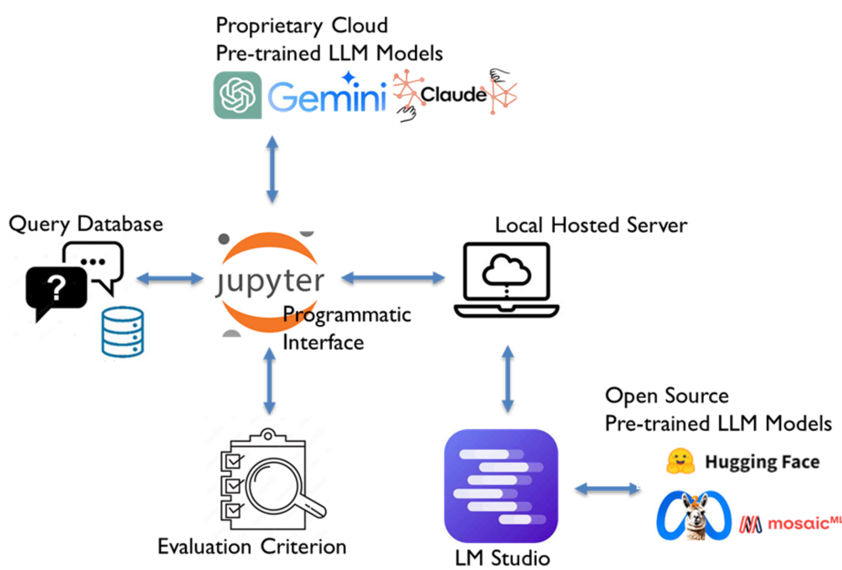


Figure 14. LLM Evaluation Framework

The code structure provided to the LLM of interest is below using LangChain. A zero shot method was used in the evaluation.

```python
import os
import json

results_dict = {}
file_output_name = 'llm_output.json'
file_path = os.path.join(folder_directory, file_output_name)
# Loop through each row in the DataFrame
instructions = "You are taking a multiple choice test. Only provide your letter answer, no explanation."
for index, row in df.iterrows():
    print("\nQuestion #" + str(row['Q#']) )
    prompt = f"Question: {row['Question']}\nChoices:\nA. {row['Choice A']}\nB. {row['Choice B']}\nC. {row['Choice C']}\nD.
    print(prompt)
    result = chat([SystemMessage(content=instructions), HumanMessage(content=prompt)])
    print(result.content)
    # Use the question number (or any unique identifier) as the key
    results_dict[row['Q#']] = result.content

    # Write the results_dict to a JSON file in the specified directory
    with open(file_path, 'w') as file:
        json.dump(results_dict, file)
```

Figure 15. Query and Response Code

Each language model's responses are to be scored against the correct answer key. The final performance of a model against the benchmark is to be represented as a percentage correct.

Results, Discussion, and Limitations

Results and Discussion

The implementation of SysEngBench across a range of LLMs, including both quantized models for local deployment yields insightful results into the capabilities and limitations of current AI technologies in the context of systems engineering. This section presents a summary of the findings, drawing comparisons between model performances and discussing the implications for the application of LLMs in systems engineering. The results for running the current state of the benchmark through open source LLMs is below in **Figure 25**.

			0.784482759	0.896551724	0.793103448
Row Labels	**Question Count**	**Question %**	**20240320 LLaMA 2**	**20240320 Mistral**	**20240320 Orca 2**
Fundamentals of SE	**116**	**100.00%**	**91**	**104**	**92**
SE Definitions	9	7.76%	8	9	9
Problem Definition and Stakeholders	11	9.48%	7	8	7
MBSE Overview	4	3.45%	3	3	3
Requirements	22	18.97%	17	22	15
Functional Analysis	11	9.48%	6	6	5
Value System Design	13	11.21%	12	12	13
Architecture	6	5.17%	3	5	4
Decision Making	10	8.62%	7	9	7
Risk	3	2.59%	2	3	3
System Integration, Qualification, Costs, Life Cycle Issues	27	23.28%	26	27	26
Grand Total	**116**	**100.00%**	**91**	**104**	**92**

Figure 16. Benchmark Evaluation Results

Out of the three models tested, the best performing model was Mistral at 89%, followed by Orca 2 at 79%, and Llama 2 at 78%. Perhaps the biggest delineating factor was performance of the models with Requirements questions, where Mistral was a clear leader with 22 correct out of 22, followed by Llama 2 with 17 and Orca with 15.

The worst performing topic for Llama 2 by percentage was architecture, for Mistral by percentage was functional analysis, and for Orca 2 by percentage was functional analysis as well. Should this trend continue, RAG or fine-tuning for functional analysis would be a potential knowledge gap solution, although not enough data points currently exist in the benchmark to statistically determine detrimental performance for the subtopics within systems engineering.

Challenges and Limitations

During the benchmarking process, a few challenges arose:

1. Very few LLM answers would have a letter selection followed by the choice verbiage and/or justification
2. Iterative refinement of the system message was required until the output was constant

Going forward, tighter integration with LangChain and lm-evaluation-harness should solve these issues.

The presence of variance by shifting which letter has the correct answer has been studied and is known to be present (Zheng et al., 2024). The variance for correct answer letter selection has not yet been investigated for this dataset.

The variance in different levels of quantization for different models was not tested. Open source versus proprietary was not yet tested, although the framework will allow for such an analysis in future work.

The level of complexity of questions within SysEngBench was also at a low complexity, focusing on high level concepts, and lacked a plethora of specific case studies.

Implications and Future Work

Implications for Systems Engineering

The SysEngBench benchmark has provided initial insight into capabilities and limitations of Large Language Models (LLMs) within the field of systems engineering. As the benchmark continues to be developed and LLMs progress over time, SysEngBench will allow for a reliable baseline for understanding model performance in systems engineering.

Eventual implications include enhanced efficiency and reduction of cognitive load required for tasks like documentation review, compliance checks, and stakeholder communications, enabling engineers to focus more on higher level aspects and navigating the available trade space of the complex system.

Future Directions and Related Work

The results of SysEngBench should be interpreted with consideration of its limitations, including the scope of tasks and the inherently complex nature of systems engineering characterized by the presence of multiple viable solutions.

Future iterations of the benchmark will incorporate a wider range of tasks, improved metrics for evaluating creative and integrative thinking, and direct comparisons with human performance to further refine our understanding of LLMs' potential in systems engineering. Various levels of complex questions, derived from a mix of real-world case studies, expertly crafted synthetic scenarios, and annotated datasets from academic and industry sources will be paramount.

A comprehensive list of future benchmark enhancements and research directions:

- **Complex Question Expansion**: To further challenge LLMs and accurately gauge their proficiency, SysEngBench will incorporate a broader array of complex questions and case studies that demand higher-order thinking, problem-solving, and the application of deep domain-specific knowledge. This expansion aims to push the boundaries of what LLMs can achieve within systems engineering.
- **Subfield Diversification**: Future iterations of the benchmark will expand upon the topic areas, such as safety engineering, logistics, sustainability, and human factors

engineering. This diversification will ensure that SysEngBench more fully represents the interdisciplinary nature of systems engineering and its varied applications across industries.

- **Evaluation by Practicing Systems Engineers**: Establish a comparative baseline and validate the benchmark's relevance; SysEngBench will be administered to practicing systems engineers. This initiative seeks to benchmark human performance against that of LLMs, offering invaluable insights into areas where AI can complement human expertise and identifying gaps where further AI development or human oversight is required.
- **Evaluation of Multiple Choice Question Bias within SysEngBench**: Evaluate the bias within multiple models across all topic areas to determine the variance of choosing correct answers. Leverage the techniques performed by Zheng et al. (2024).
- **Multimodal Input and Output**: Incorporate multimodal inputs (e.g., diagrams, charts, and technical drawings) and evaluating models' abilities to generate multimodal outputs could enhance the relevance and applicability of the benchmark to systems engineering practices.
- **Systems Engineering Domain Specific LLMs**: Investigate approaches to customize or specialize LLMs for specific domains within systems engineering via RAG or fine-tuning. Compare domain specific performance against the SysEngBench.
- **Enabling Round Table AI Discussions for an LLM SE Team:** Create a simulated team where multiple LLMs, each specialized in different aspects of systems engineering, can interact in a roundtable discussion format to tackle complex engineering challenges. The goal is to assess how well these AI models can collaborate, share insights, and come to a consensus or offer a range of solutions when confronted with multifaceted systems engineering problems.

Some of the future directions above include collaborations with others within the research group that are also working on the following topics:

- **Small Language Models for Domain Specific Knowledge:** Unlike their larger counterparts, these models aim to achieve deep expertise in narrow areas, potentially offering more precise and nuanced understanding and solutions. This approach could significantly enhance the quality of AI-generated recommendations and analyses in specialized fields, making these models invaluable tools for experts requiring detailed, domain-specific information.
- **Evaluation of LLMs with SysMLv2 Queries:** Evaluating LLMs' ability to understand and generate SysMLv2 queries represents a critical step towards integrating AI more deeply into systems engineering workflows. Current research investigates LLMs on their capacity to parse, reason about, and manipulate SysMLv2-based models, potentially automating or augmenting aspects of the systems modeling process (Longshore et al., in press). Success in this area could accelerate the model-based systems engineering (MBSE) process, making it more efficient and accessible.
- **Evaluation of LLMs for Modern Systems Engineering Cost Modeling with COSYSMO:** Constructive Systems Engineering Cost Model (COSYSMO) represents a cornerstone for estimating the costs associated with systems engineering projects. By evaluating LLMs on their ability to apply COSYSMO principles and methodologies, research can uncover AI's potential to revolutionize cost estimation in systems engineering in addition to accounting for AI productivity in novel cost factor modeling

(Madachy et al., in press). LLMs could assist in dynamically adjusting cost models based on real-time data and trends, offering a more agile and accurate approach to project management and budgeting in the field of systems engineering.

Conclusion

SysEngBench represents a significant advancement in evaluating the potential of Large Language Models within systems engineering, illuminating both the current capabilities and future possibilities of AI. By expanding the benchmark to encompass more intricate questions, a wider array of systems engineering subfields, and incorporating evaluations by practicing engineers, SysEngBench aims to bridge the gap between theoretical AI performance and practical engineering expertise. The evolving symbiotic relationship between AI development and systems engineering practice not only augments the capabilities of Large Language Models (LLMs) but also heralds a new era of engineering innovation characterized by collaborative partnerships between humans and AI. As we continue to explore the frontier of AI in systems engineering, the insights gained from SysEngBench will undoubtedly play a crucial role in shaping the future and maturing the discipline of systems engineering.

References

AI fundamentals: Benchmarks 101. (2023, April 7). https://open.spotify.com/episode/16vo3YLUTZI0nwAbrhjWYT

Boehm, B. (1986). *A spiral model of software development and enhancement*.

Clark, P., Cowhey, I., Etzioni, O., Khot, T., Sabharwal, A., Schoenick, C., & Tafjord, O. (2018). *Think you have solved question answering? Try ARC, the AI2 reasoning challenge* (arXiv:1803.05457; Version 1). arXiv. http://arxiv.org/abs/1803.05457

Defense Acquisition University. (2024). *Model-based systems engineering*. https://www.dau.edu/datl/b/model-based-systems-engineering

Dua, D., Wang, Y., Dasigi, P., Stanovsky, G., Singh, S., & Gardner, M. (2019). *DROP: A reading comprehension benchmark requiring discrete reasoning over paragraphs* (arXiv:1903.00161; Version 2). arXiv. http://arxiv.org/abs/1903.00161

EleutherAI/lm-evaluation-harness. (2024). [Python]. EleutherAI. https://github.com/EleutherAI/lm-evaluation-harness (Original work published 2020)

Google/BIG-bench. (2024). [Python]. Google. https://github.com/google/BIG-bench (Original work published 2021)

Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., & Steinhardt, J. (2021). *Measuring massive multitask language understanding* (arXiv:2009.03300; Version 3). arXiv. http://arxiv.org/abs/2009.03300

HuggingFace. (2024a). *Open LLM leaderboard*. https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard

HuggingFace. (2024b). *Quantization*. https://huggingface.co/docs/optimum/en/concept_guides/quantization

Lab42. (2024). *About ARC*. https://lab42.global/arc/

Lambert, N. (2023, November 24). *Synthetic data: Anthropic's CAI, scaling, OpenAI's superalignment, tips, and open-source examples*. https://www.interconnects.ai/p/llm-synthetic-data

Llama 2: Open foundation and fine-tuned chat models. (2024). https://ai.meta.com/research/publications/llama-2-open-foundation-and-fine-tuned-chat-models/

Longshore, R., Madachy, R., & Bell, R. (in press). *Leveraging generative AI to create, modify, and query MBSE Models*. 21st Annual Acquisition Research Symposium.

Madachy, R., Bell, R., & Longshore, R. (in press). *Systems acquisition cost modeling initiative for AI assistance*. 21st Annual Acquisition Research Symposium.

Packt. (2024). *Generating synthetic data with LLMs*. https://www.packtpub.com/article-hub/generating-synthetic-data-with-llms

Paperno, D., Kruszewski, G., Lazaridou, A., Pham, Q. N., Bernardi, R., Pezzelle, S. … Fernández, R. (2016). *The LAMBADA dataset: Word prediction requiring a broad discourse context* (arXiv:1606.06031). arXiv. https://doi.org/10.48550/arXiv.1606.06031

SEBoK. (2024). https://sebokwiki.org/wiki/Guide_to_the_Systems_Engineering_Body_of_Knowledge_(SEBoK)

Synthetic data: Save money, time and carbon with open source. (2024). https://huggingface.co/blog/synthetic-data-save-costs

Talamadupula, K. (2024, February 21). *A guide to quantization in LLMs*. Symbl.Ai. https://symbl.ai/developers/blog/a-guide-to-quantization-in-llms/

Vaneman (Director). (2020, March 31). *Webinar: Model-based systems engineering de-mystified*. INCOSE YouTube. https://www.youtube.com/watch?v=BPlphC88xR4

Zellers, R., Bisk, Y., Schwartz, R., & Choi, Y. (2018). *SWAG: A large-scale adversarial dataset for grounded commonsense inference* (arXiv:1808.05326; Version 1). arXiv. http://arxiv.org/abs/1808.05326

Zellers, R., Holtzman, A., Bisk, Y., Farhadi, A., & Choi, Y. (2019). *HellaSwag: Can a machine really finish your sentence?* (arXiv:1905.07830; Version 1). arXiv. http://arxiv.org/abs/1905.07830

Zheng, C., Zhou, H., Meng, F., Zhou, J., & Huang, M. (2024). *Large language models are not robust multiple choice selectors*. The Twelfth International Conference on Learning Representations. https://openreview.net/forum?id=shr9PXz7T0

PANEL 21. ACQUIRING, DEVELOPING, AND MODERNIZING SOFTWARE-BASED SYSTEMS

Thursday, May 9, 2024	
2:15 p.m. – 3:30 p.m.	**Chair: Mark E. Krzysko,** Principal Deputy Director, Enterprise Information, Acquisition Data and Analytics ***Innovation in Software Acquisition: The Good, Bad, and Ugly*** Jeff Dunlap, Naval Postgraduate School ***DOD Current and Planned Software Modernization Efforts*** Andrew Burton, GAO

Mark E. Krzysko—is currently serving as the Principal Deputy Director for Acquisition Data and Analytics while also leading Enterprise Information. In this senior leadership role, Mr. Krzysko directs acquisition data governance, data access, and data science to enable the Department of Defense to make sound business decisions with data. His efforts have ignited a philosophical and technical transformation within the Department by advancing innovative, data-driven approaches across all 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.

Previously, Mr. Krzysko served as a Department of Defense representative to 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 was a critical stakeholder in the National Academies of Sciences, Engineering, and Medicine's Data Science Post-Secondary Education Roundtable discussion on data science education and practice, the needs of the community and employers, and ways to move forward.

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.

He began his senior executive career in the Defense Procurement and Acquisition Policy Office as the first Deputy Director of e-Business. As the focal point for the acquisition domain, he was responsible for the 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.

Before joining OSD, Mr. Krzysko led the Electronic Commerce Solutions Directorate for the Naval Air Systems Command, 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, and various senior executive positions within the retail industry.

Mr. Krzysko holds a Bachelor of Science Degree in Finance and Master of General Administration, Financial Management, from the University of Maryland University College, and numerous certificates from Harvard's Kennedy Business Schools, Pennsylvania State University, and University of Virginia Darden School of Business.

Innovation in Software Acquisition: The Good, Bad, and Ugly

Jeffrey Dunlap, CAPT USN (Ret.)—joined the Department of Defense Management at Naval Postgraduate School in 2019. Before that, he was in the Defense Industrial Base for 8 years after retiring in 2011 from military service. While in the military, CAPT Dunlap was an Acquisition Professional and the Program Manager for several software-intensive systems in development. Operationally, he was a major department head on LHA-1 and CVN-76 during major software upgrades and installations. Dunlap holds master's and bachelor's degrees in engineering. [Jeffrey.dunlap@nps.edu]

Abstract

The Department of Defense (DoD) has recognized that interconnected warfighting systems are vulnerable due to their inability to swiftly adapt to new technologies and effectively combat advanced cyber threats. The commercial sector has developed methods to rapidly and securely implement new software capabilities without significant interference to current operations. Transitioning from entrenched practices, such as exhaustive requirement reviews and protracted capability deliveries, to a more iterative and continuous deployment model poses challenges. Value Stream Management has emerged as a means to identify and address inefficiencies, such as silos and bottlenecks, that hinder the prompt delivery of capabilities to the edge of friction. The initial step towards fostering a culture of innovation and enabling the successful flow of capabilities is identifying and eliminating unnecessary delays and legacy obstacles. A coordinated effort within the DoD is necessary to ensure successful innovation in software acquisition. This effort must include identifying and modifying counterproductive organizational behaviors, empowering lean practices, and employing adaptive change management to increase delivery velocity.

Executive Summary

The Department of Defense Instruction (DoDI) 5000.87 introduces a modern Software Acquisition Pathway (SWP) to simplify the acquisition of software-centric applications. However, the Department of Defense (DoD) still faces obstacles in fostering innovation and achieving the desired outcomes within the middle tier and major capability acquisition pathways due to entrenched business practices (slow and bureaucratic) and remnants of legacy certification processes. There can be resistance to change within any large organization, and the DoD is no exception. Cultural inertia can impede modernization initiatives and complicate the adoption of novel software processes and architecture. Many DoD software-intensive systems are decades old and cannot respond to rapidly changing threats. Legacy systems can be difficult to update or replace, leading to increased costs and reduced flexibility. These challenges restrict the swift delivery of capabilities to the point of friction in a relevant time frame.

The Challenger

In January 2023, the combat systems aboard the USS Gravely failed to intercept an incoming anti-ship cruise missile in the Red Sea. This incident was a stark reminder of the unnecessary risks our sailors face due to legacy software and outdated processes. These systems, designed to counter Cold War–era threats, are hampered by their monolithic architectures and industrial-age risk-averse mindsets, making it difficult to swiftly deploy software fixes and enhancements. This failure underscores the urgent need for substantial innovation in our software acquisition programs.

The Department of Defense (DoD) has traditionally treated software acquisition as a secondary concern to hardware-centric systems. However, the Defense Science Board (DSB) and the Defense Innovation Board (DIB) have consistently voiced their concerns over the DoD's outdated acquisition processes and delays in delivering software capabilities to the warfighter. Their recommendations, backed by their extensive knowledge and experience, have repeatedly stressed the need for significant reforms to keep pace with the rapidly evolving threats and technological advancements.

Over the past 4 decades, numerous reports and studies have highlighted the DoD's reluctance to adopt new software development practices. It can be challenging to embrace change when past acquisition efforts have proven successful against threats from the Soviet era. The need for the DoD to quickly adapt to changing mission requirements is directly linked to China's influence and capabilities. Historically, the DoD has been the leader in innovative technologies, but investment by the private sector to field new capabilities with velocity has changed the playing field. The DoD can no longer control the narrative of which countries can access commercial technologies. Realizing that countries or non-state actors are accelerating their capabilities using commercial technologies and practices has jolted the DoD into action.

Resistance

Newton's laws of motion are the foundation of classical mechanics, studying how objects move and interact. As a thought exercise, with a comparison to Newton's laws, think of the DoD acquisition system, its regulations, and its behaviors as "the body or mass":

- **Inertia:** The 1st law states that a body moving at constant speed in a straight line will keep moving in a straight line unless a force acts upon it.
- **Change:** The 2nd law states that the time rate of change of a body's momentum is equal in magnitude and direction to the force imposed on it.
- **Resistance:** The 3rd law states that when two bodies interact, they apply forces to one another that are equal in magnitude and opposite in direction (action and reaction).

Using the 1st law as a general principle of inertia, individuals, teams, or organizations will operate in their current manner without an impetus to change. Assuming that change will happen within the DoD by simply willing it to happen or writing a policy without a push or a pull force is folly without clearing away obstacles, giving maneuver space and top cover.

The 2nd law of change (Force = mass × acceleration) is that the vector sum of forces (F) on an object is equal to the mass of that object multiplied by the acceleration of the object. In this comparison, mass is the DoD acquisition process bureaucracy, and a significant power influencer force (F) must be applied to effect the acceleration, to make change ($a = F/m$). Change is possible and timely when more influential forces drive the change (bigger F). A significant mistake is believing that change can come quickly from a small group or lower-echelon organization acting on the total bureaucracy (smaller F).

Newton's 3rd law could be a comparison for change resistance within a large organization, and it is also worth considering. Starting with the 1st law, we have the DoD acquisition process operating with significant built-up years of "inertia" (this is how it has always been done philosophically). As discussed in the 2nd law, a force must be applied to make change happen—and the larger, the better to have a significant impact. The 3rd law suggests that resistance to change will occur at all levels within the organization and push back on the force of change for various reasons: lack of understanding or urgency, career risk avoidance, entrenched culture or behaviors, lack of knowledge on what needs to change, or minimum training on how to make the change. The first reaction is to apply additional force to overcome the resistance, but throwing a ball hard against a wall only makes it come back faster and harder until the thrower can no longer catch the ball. Aggressively enacting change within large bureaucratic organizations has consequences. Effective change requires significant time investment to break down the causes of resistance and solve them individually. Leadership plays a crucial role as the force for change, but the time needed often exceeds political leadership appointment periods.

The Empire Force

The institutional way of developing software within DoD acquisition was primarily the waterfall method (1970s), which values completeness in requirements and design over the speed of capability delivery. Because there was no near-peer threat after the fall of the former Soviet Union in the mid-1980s, there seemed little interest in the DoD acquisition community in changing business practices to deliver quality software faster.

Meanwhile, personal computer prices became within reach of the general population, and software applications exploded with the dawn of the Internet in the 1990s, creating an "arms race" for competitive software advantage and market share. Software development methods began to evolve within the commercial software industries, allowing first-to-market strategies to emerge.

The waterfall method had significant disadvantages in delivering software quickly, and changing to a more responsive software development model was a live-or-die decision within this new commercial environment. The C suite mandated the need for change, and resistance was addressed through training and implementing different cultural values needed for the workforce. In 2001, the commercial sector described the agile software method and the behaviors required to succeed in the changing software environment.

The Rising

Iterative software methods like agile have emerged to deliver quicker initial value in smaller chunks instead of complete software packages. Potential software users now have a voice and freely give feedback to the developer, which ultimately reduces wasted effort and accelerated capability delivery. Commercial software companies recognize that having skilled software coders and developers who can innovate is a strategically important asset. They understand that fostering a workforce capable of continuously creating, improving, and delivering cutting-edge software solutions is crucial for maintaining a competitive edge in the rapidly evolving technology landscape.

As the Internet of Things exploded, so did communities of bad actors looking to exploit software and design vulnerabilities for personal or political objectives. Unfortunately, manual software testing approaches place a heavy reliance on the skills and diligence of individual testers to identify defects and issues, often towards the latter stages of the software delivery pipeline. The need to deliver quality software with speed gave rise to automated testing tools in development, which, in turn, started the philosophy of DevOps in 2007.

The Practice

DevOps practice promotes better communications and collaboration between development and operations teams to address change challenges. The term "DevSecOps" emerged around 2012, emphasizing the integration of security practices and mindset as a focal point within the software development and operations lifecycle. This evolution to DevSecOps recognized security as everyone's responsibility and that addressing security considerations early and continuously throughout the process was crucial for delivering secure, high-quality software at scale.

The DoD's recognition of China as a rising threat to national security underscores the gravity of the situation and the need to reenergize innovation. The DoD's acquisition process was seen as inadequate in responding to emergent threats (1st law), highlighting the urgent need for change in the DoD's software acquisition practices. This recognition has sent shockwaves throughout the DoD, prompting a significant rebranding of the Defense Acquisition System (DoD Directive 5000.01) in 2020.

Major Force

A major force (2nd law) for change to the bureaucracy of DoD acquisition culminated in January 2020 with the *Operation of the Adaptive Acquisition Framework* (DoD Instruction 5000.02), signed by the Under Secretary of Defense for Acquisition and Sustainment. The Adaptive Acquisition Framework (AAF) defines the influencer forces needed to effect change and empower leadership to employ thoughtful, innovative, and disciplined approaches to deliver capability to the warfighter that is relevant and timely to the fight. As part of the change, the recognition occurred that software differs from hardware systems. The Software Pathway (SWP) of Acquisition aims to facilitate rapid and iterative delivery of software capability (e.g., software-intensive systems or software-intensive components or subsystems) to the user. Additional force multipliers added responsibilities in the *Operation of the Software Acquisition Pathway* (DoD Instruction 5000.87) to institutionalize recent changes. Policy within SWP aimed to reduce the resistance to change by removing institutional barriers and insisting on demonstrating viability within a short period (3rd law).

Whiplash Effect

In DoD acquisition, waterfall software development methodologies have almost overnight become a faux pas, as the AAF SWP requires iterative software methods. DoD acquisition also became aware of the DevOps movement in the commercial space, where capabilities could be developed and delivered within days or even hours. Constantly chasing after the latest technological trend or innovation to compensate for years of neglect and stagnation can have a detrimental "whiplash effect" within an organization. This approach often leads to a lack of focus, wasted resources, and disruption to existing processes and workflows.

Recognizing that not all software systems are created equal within the DoD acquisition landscape is crucial. The diverse range of mission requirements and operational contexts necessitates considering different software development and delivery models tailored to the specific mission model at hand. A one-size-fits-all approach to software acquisition and development needs to be revised to address the varying needs and constraints of different DoD programs and systems. Due diligence requires evaluating system criticality, operational environments, security requirements, and integration with existing infrastructure to determine the most appropriate software development methodology. For instance, mission-critical systems with stringent safety and security requirements, such as those used in weapons systems or command and control applications, may necessitate a more structured and rigorous development approach, such as waterfall.

Maybe Not So Evil?

Waterfall emphasizes extensive up-front planning, documentation, and thorough testing to mitigate risks and ensure compliance with strict standards. Waterfall software development is ideally suited when all mission requirements are known, documented, and quantified. Waterfall often results in a longer development cycle and delayed feedback by design since the entire capability is delivered in one shot. Any changes or course corrections required due to evolving requirements or unforeseen issues necessitate rework and can significantly impact the delivery timeline. DoD program managers need the flexibility to decide which software development method meets the program's mission needs. Delivering partial or iterative working software capabilities is not necessarily the best option in a Tomahawk cruise missile guidance system or the engine control software in an F-35.

Sweet Spot

Agile and iterative software development methods embrace an adaptive approach, breaking the project into smaller, manageable iterations or sprints to deliver working software of value to users sooner. Frequent feedback loops and opportunities exist for course correction based on user input and evolving needs. Retiring technical debt earlier and delivering working software sooner allows for the early identification and mitigation of issues. Delivering iterative software capability based on a Capability Needs Statement for a traditional Information Technology (IT) networked or Command, Control, Communications, Information (C4I) system is squarely in the sweet spot for agile software development. Major capability and middle tier acquisitions have all had challenges implementing modern software development methods within the hardware-dominated pathway.

Agile–Scrum–Fall (BS)

Several DoD acquisition programs have taken a conservative approach to migrate from waterfall to an iterative/agile-like software development environment ("agile–scrum–fall"). At the onset, this might seem like an excellent way to balance the strengths of both methodologies. However, it often leads to more challenges and inefficiencies than a pure agile or waterfall approach. Agile and waterfall are fundamentally different in their philosophies. Trying to combine these two can lead to confusion and conflict.

One of the core principles of agile is the continuous and incremental delivery of value. This approach allows for regular feedback and adjustments, ensuring the end product is closely aligned with user needs and expectations. The waterfall methodology is designed to deliver all value at the end of the project, following a strict sequence of phases. Combining these two methodologies into an "agile waterfall" approach can lead to inefficient use of resources. The continuous and adaptive nature of agile can be hindered by the sequential structure of waterfall, potentially delaying the delivery of value. This delay can result in longer lead times, increased costs, and a final product that may not fully meet user needs due to the lack of regular feedback and adjustments. Managing an "agile waterfall" project can be more complex than managing a purely agile or purely waterfall project, as it requires balancing both methodologies' conflicting demands and processes. There is a risk that instead of getting the best of both worlds, an "agile waterfall" approach might end up with a poor implementation of both methodologies, leading to suboptimal results. The DIB report of 2019 introduced the term "agile BS" to describe ineffective implementations of agile methodologies.

Poor implementation of any software method can lead to bugs, security vulnerabilities, system instability, and a poor user experience. It can also result in wasted resources, both in terms of time and money, as significant effort may be required to fix the issues caused by poor implementation. Enhanced risk of failure can negatively impact the program or product's reputation, leading to a loss of user trust and potential funding loss.

The Architect

Waterfall and agile methodologies primarily influence the software development process but can also indirectly impact the software architecture deployment method. The development methodology (waterfall or agile) can affect how the architecture evolves, especially regarding adaptability, scalability, and responsiveness to changing requirements.

- In a waterfall approach, the software architecture is typically defined upfront and remains unchanged throughout development and deployment. Monolithic software architectures are simpler to develop and easier to test, and match the change cycle of waterfall development.

- Agile methodologies, emphasizing iterative development and continuous feedback, can lead to evolving architectures like Microservices or segmented Monoliths. Microservices are more flexible in adding capabilities or making changes while deployed in the user environment but require careful management due to inherent complexities.

The Monolith

The DoD faces challenges in modernizing and delivering capabilities at a relevant speed. One of the factors contributing to this challenge is the dominant use of monolithic software architectures. This legacy approach builds software applications as a single, interconnected unit, resulting in a cohesive and unified system. However, this architecture can lead to slow and risky changes due to the tight coupling of components. Scaling and maintaining the application can also become challenging (often called "Spaghetti Code"). Additionally, the accumulation of technical debt occurs when software defects are discovered during testing or operations, with limited options for immediate correction until the next build cycle. Monolithic architecture might allow for faster initial development and deployment, but it can become increasingly complex and challenging to manage as the application grows.

The X Factor

Microservice software architectures allow modifying or updating capability without redeploying the entire software code, which is attractive to users at the edge of friction with limited bandwidth. The choice of software architecture can impact the speed of development and the level of user engagement. Microservices architecture, while potentially more complex to set up initially, can allow for faster, more independent development of different application parts and make it easier to adapt to changing user needs.

Killing the Monolith

Legacy DoD software systems based on a monolithic architecture have few options to modernize since they are typically large, complex, and tightly coupled, making them difficult to modify, scale, or update. The lack of access to the source code and intellectual property rights further complicates modernization. The vendor lock-in situation limits the DoD's ability to adapt and evolve these systems in response to changing needs and technologies, and it stifles innovation by reducing competition. As a result, the DoD often pays premium prices for incremental improvements in a spiral development process. Modern practices can help to break down large, monolithic systems into smaller, more manageable components; enable continuous and iterative development; and foster a more competitive and innovative environment:

- A "brownfield" comes from building on a "brownfield" site, where existing structures or infrastructure must be worked around or incorporated into the new design. A brownfield project means modifying or upgrading an existing software application, system, or infrastructure. Developers must work with and consider the existing code, systems, and technologies, which can limit their options and flexibility compared to a greenfield project. Brownfield projects can leverage existing investments and resources and often have a more precise scope and predictable outcomes. One common brownfield project involves modernizing or upgrading legacy systems. These are frequently older software applications or systems that have been used for a long time and may have outdated technology, limited functionality, or maintenance challenges. The brownfield project aims to update the system, improve its performance, enhance its features, and ensure its compatibility with modern technologies. Brownfield projects can also involve integrating multiple existing systems into a cohesive solution. Some advantages of brownfield efforts are making software improvements, fixing bugs, optimizing performance, and addressing security vulnerabilities.

- The Strangler Pattern is a software development approach that gradually transforms a legacy system into a new one. The name comes from the strangler fig, which grows on another plant, gradually enveloping and replacing it. The Strangler Pattern involves building a new system around the edges of the old system. The new system intercepts calls to the old system, handling those it can and passing on those it cannot to the old system. Over time, more and more functionality is moved from the old system to the new system until, eventually, the old system is "strangled"—that is, it becomes redundant and can be retired. This approach allows for incremental modernization, reducing the risks associated with big-bang replacements. It also provides continuous value delivery, as new features can be added to the new system even when the old system is retired.
- A "greenfield" event or project refers to developing code from scratch without constraints imposed by prior work. A greenfield project means creating a new software application, system, or infrastructure without considering prior work, existing systems, or legacy code. Developers can use the latest technologies, methodologies, and best practices without being constrained by compatibility with older systems or technologies. Greenfield projects can offer more freedom and creativity, but they can also come with challenges, such as the need to make many decisions from scratch and the potential for scope creep due to the lack of defined boundaries.

The Shining

The DoD is taking a significant step forward by embracing DevSecOps as the next major advancement. Programs like Kessel Run and PEO IWS X Integrated Combat Systems are early success stories, showcasing the effective implementation of DevSecOps practices using a microservice architecture. The DoD recognizes the benefits of adopting DevSecOps, including developing and deploying secure, high-quality software faster and lowering total ownership costs. The DevSecOps approach fosters a culture of collaboration and shared responsibility for security, further enhancing the overall effectiveness of software development within the DoD. DevSecOps leads to more secure, high-quality software delivered faster.

G-Forces

An illustrative comparison of g-force can describe the force exerted on DoD acquisition during rapid acceleration or deceleration within the organization. Higher g-forces, such as those experienced in high-speed maneuvers, can exert greater forces on the acquisition system, leading to increased risk and potential physiological effects of overpromising results. Adopting DevSecOps can significantly accelerate the software development process. DevSecOps (also known as Software Factory) delivers faster and more frequent releases into the production environment and becomes available to operations based on their need. The acceleration g-forces of continuous integration/continuous delivery (CI/CD) and automation principles of DevSecOps have exposed secondary challenges with the DoD's cut-and-paste adoption of commercial practices. DevSecOps integrates security and testing throughout the development lifecycle, enabling potential issues to be addressed earlier and more quickly via automation. DoD manual processes left unchecked negate the value proposition of DevSecOps. The DoD must adopt new processes, culture, and mindset rather than simply applying legacy processes to the new methodology. Shifting "left" authorizations and certifications to rapidly field new capabilities to the user often requires tremendous force (Newton's laws) beyond the program manager's capabilities.

Speed Bottlenecks

CI/CD accelerates DevSecOps, where automated builds and tests are run continuously and delivered to the production environment. Applying legacy software certification processes to DevSecOps creates a speed bottleneck. Legacy software development approaches often involve long development cycles, with integration and testing phases happening after development. Manual testing and risk-averse siloed teams can decelerate the advantages gained in the DevSecOps software development lifecycle. For example, if the Authorization to Operate is performed manually and only at the end of the development cycle (as is common in legacy processes), it can delay software delivery.

Everything Changes With Release Frequency (Be Careful What You Ask For)

The cadence of continuous software delivery (or deployment) into the operational environment dramatically impacts the DevSecOps architectural investments aligned with the need for agility. The operational "user burden" becomes the driving factor for the release frequency.

- **Major new features** that require system installation downtime and training will follow a more deliberate release and internal testing process. The AAF SWP calls for new capabilities cadence to be delivered to operations at least annually.
- **Minor updates**, including small feature tweaks and bug fixes, can also require system installation downtime. The delivery cadence depends on the software factory's architectural setup and the user windows that bring down the system. With minor process modifications, quarterly or monthly updates are possible.
- **Plugins**—new features or functionality added to existing applications without changing their code—can be updated without downtime and are the least intrusive to the user. Several weekly releases are possible but require significant process modifications within the software factory.

Pit of Despair

A world-class DevSecOps Software Factory architecture performing at one cadence is terrible at another. A yearly release cadence does not require the degree of automation, tooling, and lean processes as one on a quarterly or monthly release cadence. Shifting to a weekly or daily release cadence requires an entirely different Software Factor architecture and automated processes. Figure 1 shows how changing the cadence from monthly to weekly leaves significant gaps in the software factory operations. Everything must change to achieve a weekly or faster delivery cadence, including a new architecture, tooling, and processes.

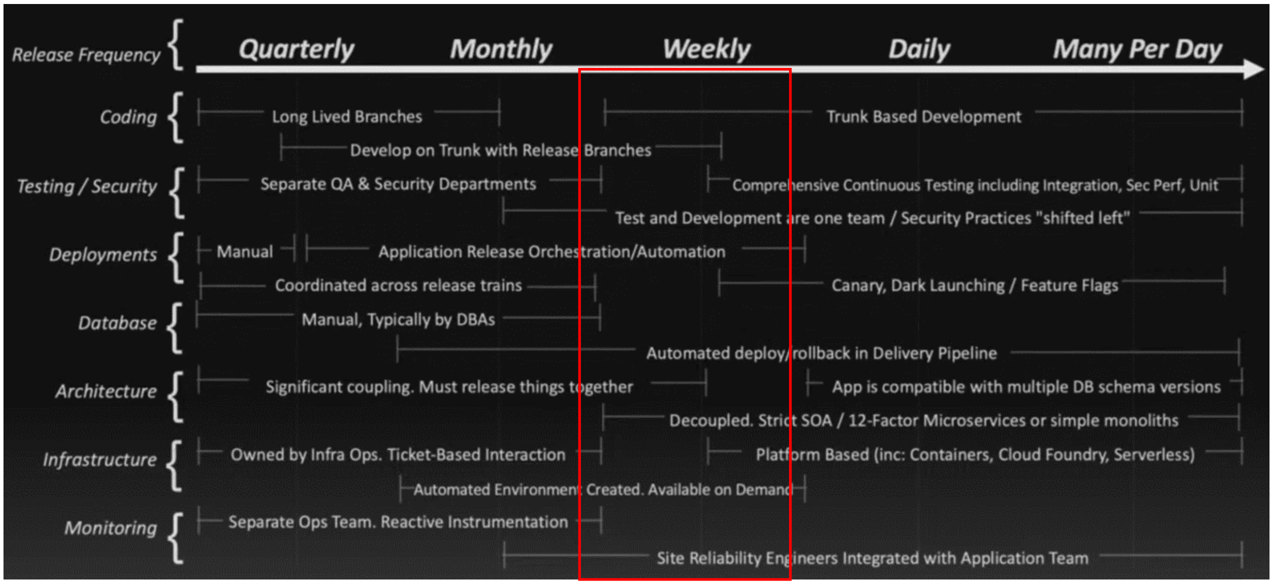


Figure 1. The Original Version Is From Lean Enterprise by Jez Humble et al.

Supporting Roles

Few software acquisition programs have the same constraints, and the AAF recommends that program managers consider the program's unique characteristics and tailor it within boundaries. The following ideals are presented as DoD acquisition pivots to employ a modern software process:

- **Top Cover (Navy Example):** Support for the software acquisition program has to span from top to bottom of the organization. The team should include supporting members to achieve synergy, including the Resource Sponsor (OPNAV), the User (Fleet), The Decision Authority (PEO), the Installation Activity and Shipbuilding Acquisition Program Management (SHAPM), Office of the Secretary of Defense, Congress, and Industry.
- **Clear and Well-Defined Requirements:** Clearly define the requirements and objectives of the software program. The first step is understanding the end users' and stakeholders' needs and any regulatory or compliance requirements. A clear understanding of the desired outcomes will help guide the development process and ensure the software meets the intended purpose.
- **Effective Project Management:** Implement effective project management practices to ensure the software program is delivered on time, within budget, and according to the defined scope. Basics like establishing a project plan, setting realistic timelines, allocating resources appropriately, and regularly monitoring progress are essential.
- **Agile Development Methodology:** Consider adopting an agile development methodology like scrum or kanban. Agile methods promote iterative development, collaboration, and flexibility, allowing continuous improvement and adaptation to changing requirements. This approach can help mitigate risks and ensure the software program meets evolving needs.
- **Skilled Development Team:** Assemble a skilled and experienced development team with the technical expertise to execute the project successfully. The government team should include software engineers, designers, testers, and other relevant roles. A competent team will be able to effectively translate requirements into functional software and address any technical challenges that may arise.
- **User-Centric Design:** Prioritize user experience and usability in the software's design. Engage with end users to gather feedback and incorporate their needs and preferences

throughout development. A user-friendly interface and intuitive functionality will enhance user adoption and satisfaction.

- **Modular Contracting:** Traditional DoD contracts often aim to deliver a complete, fully functional system at the end of the contract period. This approach can be problematic for software development, where delivering functionality in smaller increments and iterating based on feedback is often more effective. Modular contracting, which breaks down the project into smaller, more manageable pieces, has been recommended as a best practice but has faced resistance. Each capability module is contracted separately, allowing for more flexibility, risk management, and opportunities for innovation.
- **Other Transaction Authority Contracts:** Other Transaction Authority (OTA) is a vehicle federal agencies use to obtain or advance research and development (R&D) or prototypes. OTAs are not subject to the Federal Acquisition Regulation (FAR), which gives agencies more flexibility in their contracting processes. However, there are still rules and guidelines that must be followed. OTAs can be used for basic, applied, and advanced research and prototype projects. If a prototype project is successful, the agency can award a follow-on production contract or transaction without competition. At least one nontraditional defense contractor or nonprofit research institution must significantly participate in the project to use an OTA. Competitive procedures are generally used to award OTAs. Still, an agency may award a noncompetitive OTA in certain circumstances, such as when a particularly innovative concept or capability is involved or when time is of the essence.
- **Defending the Budget Using DevSecOps:** It is essential to highlight the value and benefits that DevSecOps brings to the DoD. Potential cost savings can be achieved by implementing DevSecOps practices. By integrating security and quality assurance throughout the software development lifecycle, you can identify and address issues earlier, reducing the cost of fixing them later in the development process or during production. Additionally, automation and CI/CD pipelines can streamline processes, reduce manual effort, and improve efficiency, resulting in cost savings over time. By integrating security practices from the beginning of the development process, vulnerabilities and risks can be identified and addressed early on. This proactive approach helps to minimize the potential for security breaches and reduces the associated costs and operational damage that can result from security incidents. DevSecOps promotes collaboration, communication, and shared responsibility among development, security, and operations teams. This cultural shift can improve teamwork, increase innovation, and foster a more efficient and productive work environment.
- **Ongoing Support and Maintenance:** Plan to continue software program support and maintenance after its initial deployment. The focus should include addressing bug fixes, implementing updates and enhancements, and providing technical support to users. Regular maintenance and updates will ensure that the software remains secure, reliable, and aligned with evolving needs.

Nuggets (From PEO IWS X PM Integrated Combat Systems, CAPT Phillips)

- The DoD is its own unique culture, but it is not that different.
- Deliberately build your culture.
- Know your stakeholders and constantly communicate.
- Have a robust communications plan and talking points.
- Know your market and compete where you can. (Forward progress is better than none.)
- Big bang almost always fails.
- Under-promise and over-deliver.
- Hope that the competition underestimates you because that gives you an opening.

- Change takes investment, but you will have to earn it.
- Be humble. Your first idea or version is almost always bad.
- The faster you learn, the better.
- You need to educate the entire industry on what you are doing.
- Slowly force change.
- Both an industrial and digital mindset is required to be a successful change agent in the future.

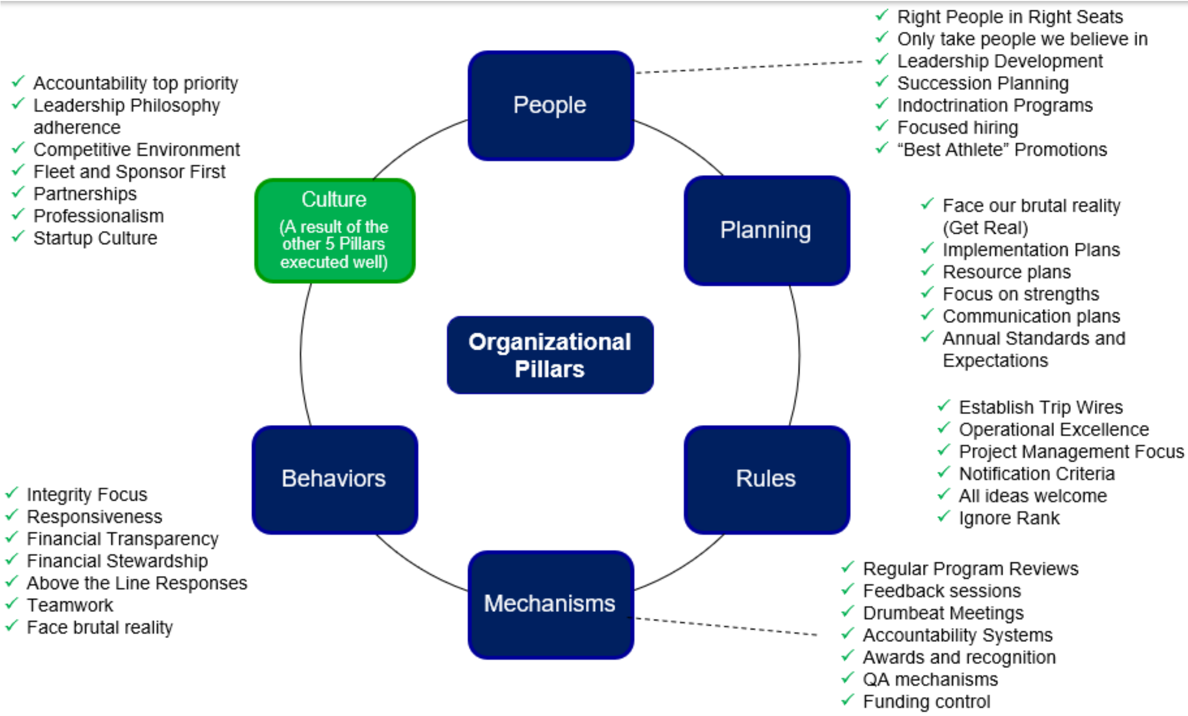


Figure 2. From PEO IWS X PM Integrated Combat Systems by CAPT Phillips

The End Game

Innovation in software acquisition within the DoD has experienced the good, the bad, and the ugly of a system accelerating without a complete understanding of the bureaucratic resistance and business practices necessary to achieve velocity.

Major considerations are needed for modular and flexible contracts, incorporating testing and evaluation throughout the software process and shifting left certifications and approvals to deploy at the speed of relevance. The importance of a trained and skilled workforce with user interaction and senior leadership support cannot be understated.

Focus is needed in understanding the significance of reasonable and prioritized requirements, advocating for a shift from compliance-based, overly prescriptive requirements to more iterative approaches like iterative/agile development approaches to reduce cost, risk, and time.

Addressing these innovation challenges requires a comprehensive approach that includes effective project management, stakeholder engagement, risk management, and a focus on iterative development and continuous improvement. Culture and behaviors take time to adjust to the applied force; they must be constant and consistent to ensure that the capability delivered is responsive to a changing threat.

DOD Current and Planned Software Modernization Efforts

Andrew Burton—is a Senior Analyst at U.S. Government Accountability Office (GAO) [BurtonA@gao.gov]

Abstract

To respond to evolving threats, DOD must develop and deliver software-based weapon and IT systems quickly. In April 2023, GAO published a report examining the extent to which DOD implemented software modernization recommendations from the Defense Science Board and Defensive Innovation Board and positioned itself to pursue future software modernization reforms. In large part, these recommendations—and DOD's planned and ongoing efforts—focused on providing innovative software-based capabilities to the warfighter by tailoring DOD's traditional processes, such as streamlining acquisition processes, employing digital transformation, piloting novel funding approaches, and providing just in time training. Drawing on recent GAO work, this presentation will focus on DOD's efforts to date to modernize how it develops and acquires software and will also offer observations on DOD's planned software modernization efforts and GAO's recommendations to improve DOD's ability to implement them.

Keywords: Software, modernization, acquisition

Background

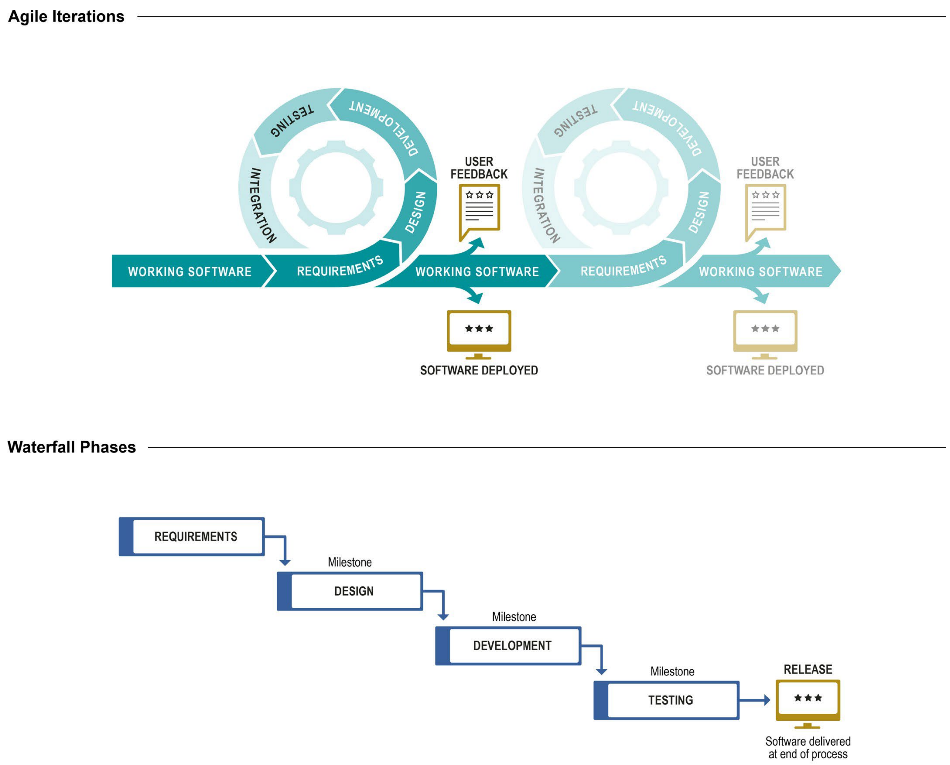
For years, commercial companies have recognized the value of software for providing new capabilities to consumers. According to the DSB and DIB, the commercial industry has developed leading practices that foster quicker, more cost effective software development, which allows for the speedier delivery of new capability to users and consumers.

DOD has also recognized software as an increasingly critical element for meeting weapon systems' requirements. However, our recent work has highlighted that DOD's software development practices have not kept up with leading industry practices even as software has become increasingly vital to DOD systems. Other recent studies, such as the 2018 DSB and 2019 DIB reports, also found deficiencies in software development and acquisition practices within DOD, such as outdated acquisition processes and delays in delivering software to users.

Agile Software Development

Modern approaches to software delivery rely extensively on Agile development. Agile development is a flexible, iterative way of developing software that delivers working capabilities to users earlier than traditional DOD software development processes, known as the waterfall approach. In most instances, adopting Agile methods involves new behaviors and a different mindset, which is a major shift in how an organization operates. For example, Agile practices call for the integration of planning, design, development, and testing into an iterative life cycle to deliver software early and often, ranging from every few days to every 60 to 90 days. The frequent iterations are intended to effectively measure progress toward delivery of the full suite of capabilities, reduce technical and programmatic risk, and be responsive to feedback from stakeholders and users.

In contrast, under the waterfall approach traditionally used by DOD, requirements are established in advance of development, and software is usually delivered as a single completed program at the end of the development cycle. Software development occurs without continual user involvement or feedback, and programs may not be able to modify requirements without cost increases and schedule delays. This software development approach mirrored the development of a DOD hardware system. Figure 1 compares Agile and waterfall approaches for developing software.



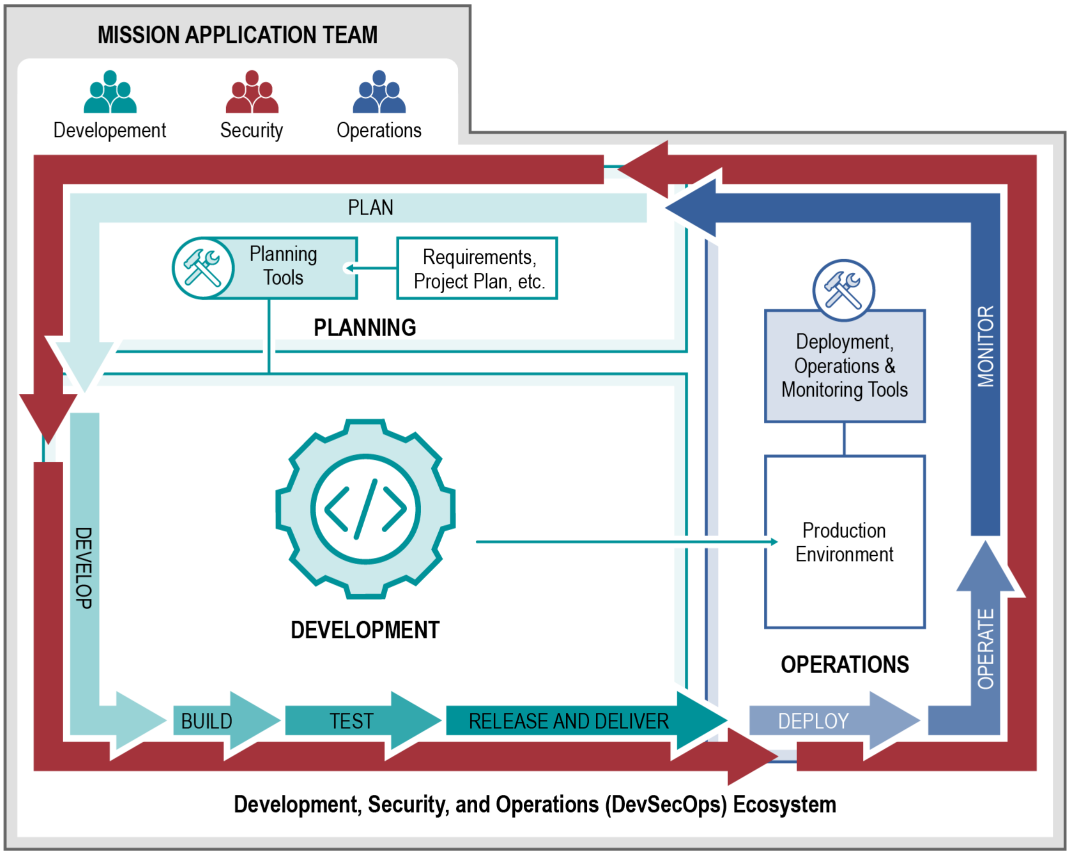
Source: GAO analysis of Department of Defense and U.S. Citizenship and Immigration Services Information. | GAO-23-105611

Figure 1. Comparison of Agile and Waterfall Frameworks for Developing Software

There are numerous frameworks available for Agile programs to use, such as Development, Security, and Operations (DevSecOps), an iterative software development methodology that combines development, security, and operations as key elements in delivering useful capability to the user of the software. These frameworks provide a basic structure to guide projects. Agile, as a concept, is not prescriptive but rather an umbrella term for a variety of iterative software approaches. Each framework is unique and may have its own terminology for processes and artifacts (documents, data, or other information describing what was planned or completed). According to GAO's Agile Assessment Guide, when implementing Agile in the federal environment, both government and contractor staff should work together to define the Agile terms and processes to be used for particular programs. The frameworks are not mutually exclusive and can be combined.

DOD's Software Factory Ecosystem

DOD's software factory ecosystem is a collection of tools and processes that support activities throughout the DevSecOps life cycle. Software factories use cloud-based computing to assemble a set of software tools enabling developers, users, and management to work together on a daily tempo. As shown in figure 2, these tools and processes support continuous iterative development through three key phases: planning, development, and operations, with security emphasized throughout each.



Source: GAO analysis of Department of Defense information. | GAO-23-105611

Figure 2. The Department of Defense's Software Factory Ecosystem

- **Planning.** This phase involves activities that help projects manage time, cost, quality, risk and other issues, such as system design, project plan creation, risk analysis, and business requirements gathering.
- **Development.** This phase contains multiple work streams, equipped with tools and workflows to automate activities with minimal human intervention to produce software applications.
- **Operations.** In this phase, software is deployed to the end user. Among other things, operations and security monitoring are performed during this time.

In February 2018, the DSB stated that software factories are a crucial part of iterative development practices, as they allow programs to identify errors and obtain user feedback continuously.

DOD's Adaptive Acquisition Framework and Software Acquisition Pathway

In January 2020, DOD reissued and updated its acquisition policies, emphasizing speed and agility in the acquisition process. The updated instruction established the Adaptive Acquisition Framework, comprised of six acquisition pathways, each tailored to the characteristics and risk profile of the capability being acquired. These six acquisition pathways are intended to, among other things, deliver solutions to the end user in a timely manner (see fig. 3).

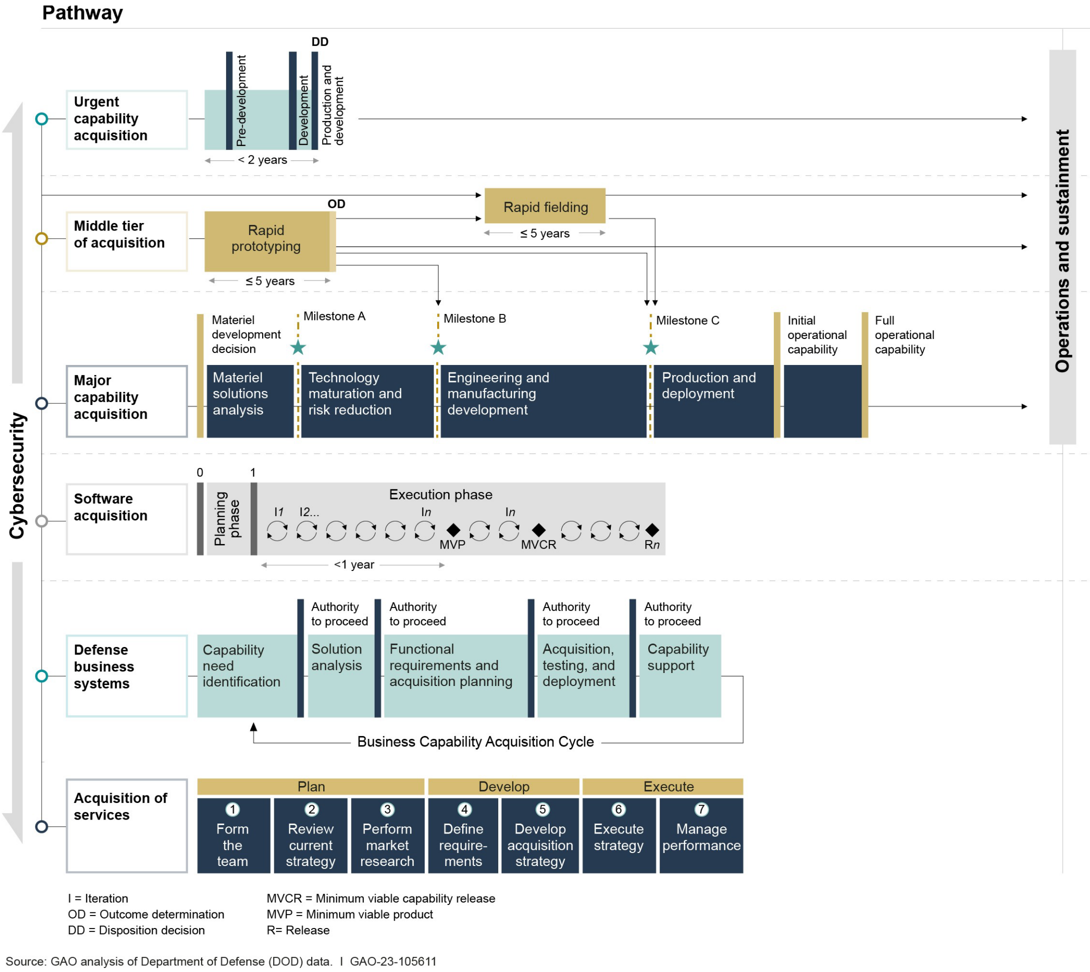


Figure 3. DOD's Adaptive Acquisition Framework

One of these pathways, the software acquisition pathway, is intended to provide for the efficient and effective acquisition, development, integration, and timely delivery of secure software. Section 800 of the NDAA for Fiscal Year 2020 mandated that DOD develop this pathway. The pathway establishes a framework for software acquisition and development investment decisions that addresses tradeoffs between capabilities, affordability, risk tolerance, and other considerations. It has two phases: planning and execution (see fig. 4).

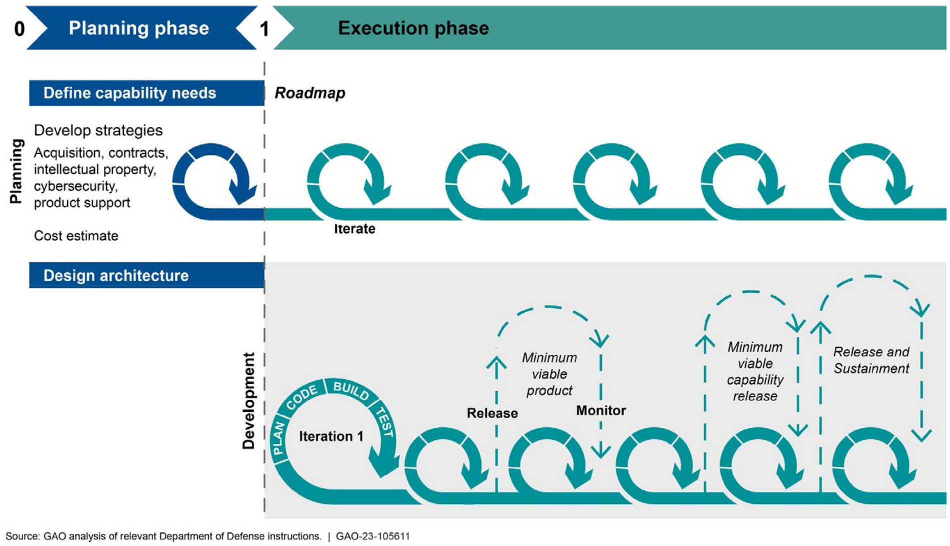


Figure 4. The Department of Defense's Software Acquisition Pathway

Using this pathway, small cross-functional teams—users, testers, software developers, and cybersecurity experts—are expected to be able to deliver software rapidly and iteratively to meet user needs. DOD policy encourages program officials to frequently engage with users and deliver new capabilities to operations at least annually. The instruction implemented recommendations we made in 2019 that DOD ensure its software development guidance provides specific, required direction on the timing, frequency, and documentation of user involvement and feedback. Further, in March 2022, we reported that the instruction generally reflected key product development principles used by leading companies.

While the software acquisition pathway offers a number of potential ways to improve DOD's ability to benefit from modern software development approaches, our recent work also shows that DOD is still determining how it will conduct oversight of the pathway. For example, we reported in June 2021 that DOD had yet to collect the data and develop tools it needed to oversee the programs using the pathway.

In September 2021, DOD stated that it had established a software acquisition pathway data collection strategy and shared it with component headquarters and relevant program offices. In addition, DOD stated that it plans to prepare a semiannual reporting template and collect trial submissions from early pathway programs to gain insights, implement suggestions, and improve the template.

Entities Involved in Software Modernization Efforts

The Under Secretary of Defense for Acquisition and Sustainment (USD(A&S)), the Under Secretary of Defense for Research and Engineering (USD(R&E)), and the DOD Chief Information Officer (CIO) are responsible for leading the coordination of software modernization activities, specifically through the Software Modernization Senior Steering Group (SSG). Among other things, the Software Modernization SSG is intended to promote the adoption of modern software development practices across the department and remove barriers to adoption.

Many other offices within OSD—including Cost Assessment and Program Evaluation (CAPE), and the Director, Operational Test and Evaluation (DOT&E)—as well as Joint Staff, and the military departments also have responsibilities for executing or overseeing certain aspects of software modernization. These organizations are also represented on the Software Modernization SSG, among others.

Examples of selected responsibilities of these offices related to software modernization include:

- USD(A&S) establishes software acquisition and sustainment policies, such as DOD's software acquisition pathway instruction.
- USD(R&E) establishes policies and advises on all aspects of defense research and engineering and technology development, such as advancing and enabling the rapid transition of software-developed capabilities to acquisition programs of record through research and development and science and technology initiatives.
- DOD CIO develops strategy and policy on the operation of DOD information technology, information systems, and cybersecurity, such as co-leading the development of DOD's Software Modernization Strategy.
- DOT&E establishes DOD testing policies, including DOD Instruction 5000.89, Test and Evaluation, which outlines testing guidance for software acquisition pathway programs.
- CAPE establishes policy on cost estimation and analysis, including DOD Instruction 5000.73, Cost Analysis Guidance and Procedures, which outlines cost estimation guidance for software acquisition pathway programs.
- Joint Staff develops supplemental guidance for requirements validation and reviews software programs for joint requirements.
- Military departments implement DOD software acquisition policy and, acting through the decision authority, oversee software acquisition pathway programs. In addition, military departments develop supplemental software policies and manage their software workforce.

DSB and DIB Software Modernization Recommendations

Established in 1956, the DSB serves as the Federal Advisory Committee chartered to provide DOD leadership with independent advice and recommendations on science, technology, and acquisition processes, among other things. The DSB is comprised of former senior military and government officials as well as leaders from academia and industry.

In February 2018, a DSB task force concluded that DOD can, and should, leverage commercial software development leading practices to its advantage, including on its weapon systems. The DSB study made seven recommendations to DOD. We reported previously that DOD was taking steps to address some of these recommendations. Table 1 provides a list of the seven DSB recommendation topics and the specific recommended actions.

Table 1: GAO Summary of February 2018 DSB Software Modernization Recommendations

Recommendation	Recommended actions
Software factory[a]	• Establish a common list of source selection criteria for evaluating software factories for use throughout the department • Require contractors to demonstrate at least a pass-fail ability to construct a software factory • Review and update source selection criteria every 5 years
Continuous iterative development[b]	• Adopt continuous iterative development best practices for software, including security, throughout the acquisition life cycle • Identify minimum viable product approaches • Delegate acquisition authority to program managers • Require all programs entering system development (Milestone B) to implement iterative processes for acquisition category I, II, and III programs • Identify best practices and incorporate into regular program reviews

Adoption of risk reduction metrics for new programs	• Allow multiple vendors to begin work. After a vendor has demonstrated that work can be done, a down-select should happen. Retain several vendors through development to reduce risk, as feasible • Modernize cost and schedule estimates and measurements and contract with the defense industrial base for work breakdown schedule data to include, among others, staff, cost, and productivity • Build a program-appropriate framework for status estimation
Current and legacy programs in development, production, and sustainment	• Plan for ongoing programs to transition to a software factory and continuous iterative development processes • Require prime contractors for ongoing programs to transition to a hybrid model (i.e., hybrid approach between iterative software development and waterfall) and incorporate continuous iterative development processes into long-term sustainment plans • Make the business case for whether to transition the legacy programs for which development is complete • Provide a quarterly status update on the transition plan for programs to the Under Secretary of Defense for Acquisition and Sustainment • Brief best practices and lessons learned across the military departments from programs that have transitioned successfully to modern software development practices

Workforce	• Develop a workforce that is competent and familiar with current software development techniques
	    • Military departments should acquire or access a small cadre of software systems architects with a deep understanding of iterative development
	    • Services acquisition commands should use this cadre early in the acquisition process to formulate acquisition strategy, develop source selection criteria, and evaluate progress
	• Develop a training curriculum, including software acquisition training, to train this cadre and ensure the program managers of software-intensive programs are knowledgeable about software
	• Direct the Defense Acquisition University to establish curricula addressing modern software practices
	• Brief the Under Secretary of Defense for Acquisition and Sustainment at least annually to demonstrate contractors' progress on adopting modern software practices
	• Hire and train a cadre of modern software acquisition experts from across the military services
	• Create an iterative development integrated product team with associated training

Software sustainment	• Direct that requests for proposals and contractor selection criteria include elements of the software framework supporting the software factory, including code and document repositories and software tools • Require contractors to provide documentation, such as test files and coding, to DOD • Consider selection of contractors based on the ability of DOD to reconstitute a contractor's software framework and rebuild binaries, re-run tests, procedures, and tools against delivered software and documentation
Independent verification and validation for machine learning	• Establish research and experimentation programs around the practical use of machine learning in defense systems with efficient testing, independent verification and validation, and cybersecurity resiliency and hardening as the primary focus points • Establish a machine learning and autonomy data repository and exchange to collect and share necessary data from and for the deployment of machine learning and autonomy • Create and establish a methodology and best practices for the construction, validation, and deployment of machine learning systems

Source: GAO analysis of Defense Science Board (DSB) information. | GAO-23-105611

DIB Recommendations

Established in 2016 under the Federal Advisory Committee Act, the DIB provides independent recommendations to the Secretary of Defense and other senior DOD leaders on emerging technologies and innovative approaches for DOD to adopt. Topics addressed by the DIB include digital modernization, software, and artificial intelligence. The DIB is comprised of national security leaders, including from academia and the private sector.

When necessary, DOD may establish subcommittees and task forces through which the DIB provides recommendations, such as the subcommittee established to examine DOD's software acquisition and development practices. The DIB reports to the Secretary of Defense and the Deputy Secretary of Defense, who may act upon the DIB's recommendations.

In May 2019, the DIB released a report that emphasized the need for DOD to deploy software quickly, focus on continuous improvement throughout the software life cycle, and develop a workforce to follow modern software development practices. The DIB study made 10 primary recommendations to address statutory, regulatory, and cultural hurdles DIB identified that DOD faces in modernizing its approach to software (see table 2).

Table 2: GAO Summary of May 2019 DIB Software Modernization Recommendations

Recommendation	Recommended actions
New acquisition pathway	Establish one or more new acquisition pathways for software that prioritize continuous integration and delivery of working software in a secure manner, with continuous oversight from automated analytics
New appropriation category	Create a new appropriation category for software capability delivery that allows software to be funded as a single budget item, with no separation between research, development, test and evaluation, production, and sustainment
Security considerations	Make security a first-order consideration for all software-intensive systems
Software features	Shift from the use of rigid lists of requirements for software programs to desired features and required characteristics to avoid requirements creep, overly ambitious requirements, and program delays
Digital infrastructure	Establish and maintain digital infrastructure within the Department of Defense (DOD) and the military departments that enables rapid deployment of secure software to the field, and incentivize its use by contractors
Automated testing and evaluation	Create, implement, support, and use fully automatable approaches to testing and evaluation, including security
Authorization to operate (ATO)a reciprocity	Create a mechanism for ATO reciprocity within and between programs, the military departments, and other DOD agencies to enable sharing of software platforms, components, and infrastructure, and rapid integration of capabilities
Source code access	Require access to source code, software frameworks, and development toolchains—with appropriate intellectual property rights—for DOD-specific code, enabling full security testing and rebuilding of binaries from source
Organization of development groups	Create software development units in each military department consisting of military and civilian personnel who develop and deploy software to the field using DevSecOps practicesb
Acquisition workforce and training	Expand the use of training programs for leadership and program managers that provide insight into modern software development and the authorities available to enable rapid acquisition of software

Source: GAO analysis of Defense Innovation Board (DIB) information. | GAO-23-105611

DOD's Efforts to Date At Least Partially Implement All DSB and DIB Recommendations

DOD has taken many steps to facilitate programs' ability to modernize software development and acquisition in recent years, which at least partially implemented all 17 DSB and DIB recommendations. DOD, however, has not implemented all recommended actions.

DOD officials told us that, while they are not required to implement these actions because the DSB and DIB are federal advisory boards, they expect they may implement some of them through future software modernization efforts. These officials told us that, in other cases, they have determined that implementing the recommended actions would be impractical.

DOD Has Partially Implemented Most DSB and DIB Recommendations

As shown in table 3, DOD has taken steps that partially address each of the DSB's seven recommendations but has not implemented all specific recommended actions for any of the recommendations.

Table 3: GAO Analysis of DOD Implementation of DSB Software Modernization Recommendations

GAO summary of DSB recommendations	Implementation of specifications
Evaluate software factories in source selection	◐
Adopt continuous iterative development best practices	◐
Adopt risk reduction metrics for new programs	◐
Transition current and legacy programs in development, production, and sustainment to continuous iterative development	◐
Begin workforce hiring and upskilling	◐
Review software sustainment documentation in source selection	◐
Independently verify and validate for machine learning	◐

Legend: ◐ = partially implemented.

Source: GAO analysis of Defense Science Board (DSB) report, Department of Defense (DOD) documents, and interviews with DOD officials. | GAO-23-105611

The following examples highlight actions taken by DOD that align with the DSB's recommendations as well as specific recommended actions DOD has not implemented.

Evaluate software factories in source selection. The DSB recommended several actions related to software factories, such as (1) establishing a common list of source selection criteria for evaluating software factories for use throughout DOD and (2) requiring that contractors demonstrate at least a pass-fail ability to construct a software factory to be considered minimally viable for a proposal. DOD has taken steps to address the recommended actions, but has not fully addressed them. For example, in August 2019, DOD published the Enterprise DevSecOps Reference Design, which establishes guidance for program managers on the DevSecOps ecosystem and life cycle, and applications. The reference design includes some guidance to assess agency and vendor software factories. However, use of the guidance is not required and the guidance does not address whether it should be used as criteria during source selection.

Transition current and legacy programs in development, production, and sustainment to continuous iterative development. The DSB recommended several actions related to transitioning programs to continuous iterative development. These include having ongoing development programs plan to transition to a software factory and continuous iterative development and briefing best practices and lessons learned across the military departments. DOD has taken steps to address the recommended actions. For example, DOD established policies and guidance related to continuous iterative development for programs within the software acquisition pathway, including a process for new and legacy programs to enter the pathway. DOD has also provided opportunities for programs to provide feedback and lessons learned about the adoption of modern software development practices. For instance, in February 2020, DOD published the Agile Software Acquisition Guidebook. The guidebook covers topics that programs should consider when transitioning to Agile practices as well as iterative development lessons learned from DOD's Agile pilots.

However, DOD has not implemented some of the specific recommended actions. For example, DOD officials stated that they do not intend to direct prime contractors to transition to a hybrid model and adopt continuous iterative development within current contracts, as recommended by the DSB. Officials noted, however, that they agree with the intent of the recommendation and that contractors who propose modern practices for future programs will likely be more competitive than contractors proposing a legacy model.

Begin workforce hiring and upskilling. The DSB recommended several actions related to workforce hiring and upskilling, such as establishing training curricula on modern software practices as well as acquiring and maintaining a small cadre of software systems architects with a deep understanding of iterative development. DOD has taken steps to address the recommended actions. For example, the Office of the USD(A&S) collaborated with the Defense Acquisition University (DAU) to establish training in Agile and DevSecOps methods for DOD software development and acquisition staff, including DOD leadership. In addition, the military departments have also expanded or are planning to expand training opportunities on software intensive systems and practices. For example, the Air Force Institute of Technology provides DevSecOps courses for leadership, including program managers. However, additional work remains for DOD to implement all of the specific recommended actions.

Defense Innovation Board

DOD has taken steps that fully or substantially implement four of the DIB's 10 recommendations and partially implement the remaining six recommendations (see table 4).

Table 4: GAO Analysis of DOD Implementation of DIB Software Modernization Recommendations

GAO summary of DIB recommendations	Implementation of specifications
Create a new acquisition pathway for software	●
Create a new appropriation category for software	●[a]
Prioritize security considerations	●
Shift from system requirements to software features	◐
Use digital infrastructure to enable rapid deployment	◐
Use automated testing and evaluation approaches	◐
Create Authorization to Operate reciprocity between programs, services, and DOD agencies[b]	◐
Use source code access to enable security testing	◐
Use organic development groups to develop and deploy software	◐
Provide acquisition workforce training for leadership and program managers	●

Legend: ● = fully or substantially implemented; ◐ = partially implemented.

Source: GAO analysis of Defense Innovation Board (DIB) report, Department of Defense (DOD) documents, and interviews with DOD officials. | GAO-23-105611

The following examples highlight actions taken by DOD that align with DIB's recommendations as well as specific recommended actions DOD has not implemented.

Create a new acquisition pathway for software. The DIB recommended that DOD establish one or more new acquisition pathways for software that prioritize continuous integration and delivery of working software in a secure manner, with continuous oversight from automated analytics. DOD has addressed the recommendation. For example, in response to a legislative requirement, DOD established a pathway for the timely acquisition of software capabilities by using an iterative approach to software development. DOD's policy for the software acquisition pathway provides opportunities for new and existing programs to join the pathway but does not require its use. Each program following the pathway must develop and track a set of metrics—using automated tools to the maximum extent practicable—to assess and manage, among other things, the performance, progress, speed, and quality of the software development, and the ability to meet users' needs. As of March 2023, there were 49 programs using the pathway.

Create a new appropriation category for software. The DIB recommended the creation of a new appropriation category for software capability delivery that allows software to be funded as a single budget item that could be used for the purposes of research, development, test, and evaluation (RDT&E), production, and sustainment. DOD has substantially addressed the recommendation. In December 2020, the Consolidated Appropriations Act of 2021 established the Software and Digital Technology Pilot Program. The Office of the USD(A&S), in collaboration with the Under Secretary of Defense (Comptroller) (USD(C)), engaged Congress to establish the pilot. The act provides for certain programs to use RDT&E funding appropriated in that act for procurement and sustainment activities. Traditionally, software development programs have funded RDT&E, procurement, and sustainment activities through distinct appropriation categories. This pilot is intended to provide additional funding flexibility for software programs, particularly those using modern software development methods, such as iterative testing.

DOD does not plan for the pilot to be a permanent solution to software funding issues. Rather, DOD views the pilot as an opportunity to test whether the use of a single appropriation category enables modern software development practices. DOD intends to use the pilot for several years and work with Congress to implement a long-term solution based on lessons learned from the pilot. The pilot originally included eight programs. In May 2022, DOD officials told us that Congress has not approved recent requests to include additional pilot programs. However, DOD continues to collect data on the pilot programs to understand the effect of this funding mechanism on software development programs. As explained in the Joint Explanatory Statement accompanying the Consolidated Appropriations Act, 2023, the Secretary of Defense is encouraged to refrain from submitting additional pilot programs in future budget submissions until DOD has demonstrated its ability to collect data on performance improvements resulting from the pilot program.

Use digital infrastructure to enable rapid deployment. The DIB recommended that DOD establish and maintain digital infrastructure within DOD and the military departments that enables rapid deployment of secure software to the field and incentivize its use by contractors. DOD has taken action to address the recommendation but has not fully implemented it.

DOD issued policy and guidance related to establishing and operating digital infrastructure, such as networks and software factories. For example, DOD's September 2019 Enterprise DevSecOps Reference Design provides programs with modern software development techniques that consider security and operations throughout, such as automated, iterative testing that begins earlier in the process. In addition, this guidance encourages programs to use software factories. DOD has also issued guidance related to the department's cloud infrastructure, intended to provide users and systems with secure internet access to and from unclassified cloud environments. According to DOD officials, each military department has established a cloud environment.

However, additional work remains related to establishing and maintaining digital infrastructure, as outlined in DOD's key strategy documents. For example, while not yet achieved, DOD's February 2022 Software Modernization Strategy establishes several goals:

- accelerating the DOD enterprise cloud environment;
- transitioning from disparate cloud efforts to an integrated cloud portfolio;
- establishing a DOD-wide software factory ecosystem;
- leveraging established software factories; and
- scaling the services across the department.

DOD Plans to Implement Some but Not All Remaining Recommended Actions

Officials from the Office of the USD(A&S) stated that they have addressed the intent of the recommendations from the DSB and DIB reports and do not plan to implement all of the specific recommended actions. According to DOD officials, the department is not required to implement specific actions recommended in the reports because DSB and DIB are federal advisory committees.

DOD officials told us that department-wide actions over the last several years have focused on encouraging—rather than requiring—programs to adopt modern software development and acquisition practices. Officials explained that this approach mitigates challenges with implementing the DSB and DIB recommendations that arose, in part, because older programs were less able to automate security and testing in a way that aligned with modern software development methods.

DOD officials told us they still plan to implement some specific recommended actions through their planned future software modernization efforts. For example, DOD plans additional actions to address DSB's recommendation that the military departments acquire or access a small cadre of software development professionals with a deep understanding of iterative development processes and practices.

According to officials from the Office of the USD(A&S), further planning to implement this part of the recommendation is underway in response to a provision in the NDAA for Fiscal Year 2022.

In other cases, DOD officials told us they chose not to implement the actions for specific reasons, such as the recommended actions being impractical. For example, they noted that DOD does not plan to fully implement the DSB's recommendation on transitioning programs to continuous iterative development. Specifically, DSB recommended that prime contractors—within contract constraints—transition from waterfall to a more iterative software development approach, using a hybrid approach, if necessary, and incorporate iterative development into a long- term sustainment plan. Officials from the Office of the USD(A&S) stated that they do not intend to direct contractors to take these actions because it is unrealistic to do so for a large number of contracts. These officials added that programs can make assessments of individual contracts once they have an understanding of modern software development practices.

DOD's software modernization efforts are still underway, and, moving forward, DOD is focused on continuing efforts in the areas DSB and DIB emphasized. DOD officials stated that, as the department continues its software modernization efforts, they expect that additional actions recommended by DSB and DIB will be implemented. However, these officials also noted that certain steps recommended by DIB may become outdated as time passes and technology changes.

DOD Is Not Fully Positioned to Implement Future Software Modernization Efforts

DOD has outlined planned actions to continue its software modernization efforts across the department but has yet to incorporate certain key practices our prior work shows could help DOD implement these actions successfully. While DOD's planning incorporated some elements of most of the practices we assessed, we identified gaps in the implementation of several of them.

DOD Plans Outline Transformational Future Software Modernization Efforts

The perspectives of acquisition and T&E decision-makers—*IDSK stakeholders*—form the basis for the IDSK RA viewpoints and corresponding views. A viewpoint as stated in the *Software, Systems, and Enterprise—Architecture Description ISO Standard* (ISO/IEC/IEEE, 2022) establishes the conventions for creating, interpreting, presenting, and analyzing a view to a DOD senior leadership has repeatedly emphasized the importance of ongoing software modernization efforts and the need for the department to take further actions. In a February 2022 memorandum approving the DOD Software Modernization Strategy, the Deputy Secretary of Defense stated that achieving faster delivery of software capabilities requires the combined focus of DOD senior leadership and significant changes in policies, technologies, processes, and workforce.

DOD has detailed its plans for future software modernization efforts in three key department-wide strategies.

- Digital Modernization Strategy. Published in July 2019, this strategy supports implementation of the 2018 National Defense Strategy lines of effort involving cloud, artificial intelligence, command, control and communications, as well as cybersecurity.
- Software Modernization Strategy. Published in February 2022, this strategy is one of a set of sub-strategies of the Digital Modernization Strategy. The strategy provides a framework of technologies, approaches, and processes that must be addressed to modernize software delivery, such as adoption of DevSecOps, process and policy transformation, and workforce.
- Software Science and Technology Strategy. Published in November 2021 in response to a requirement in the NDAA for Fiscal Year 2020, this strategy is intended to guide strategic thinking within DOD to advance and enable the rapid transition of software- developed capabilities to acquisition programs through research and development and science and technology initiatives. According to an official from the Office of the USD(R&E), the goals of this strategy align with the Software Modernization Strategy, but the Software Science and Technology Strategy is focused on the research and development of critical technologies while the Software Modernization Strategy aims to achieve faster delivery of software capabilities in support of DOD priorities.

Together, these strategies document the breadth of DOD's future software modernization efforts. Each plan includes a discussion of the department's vision and goals relevant to the scope of the plan (see fig. 5).

DOD Digital Modernization Strategy	DOD Software Science and Technology Strategy	DOD Software Modernization Strategy
JULY 2019	NOV. 2021	FEB. 2022
Vision Deliver a more secure, coordinated, seamless, transparent, and cost-effective IT architecture that transforms data into actionable information and ensures dependable mission execution in the face of a persistent cyber threat.	**Vision** Deliver resilient software capabilities at the speed of relevance. For instance, modernize development approaches to deliver secure, resilient software capabilities within hours or days rather than months or years.	**Vision** Deliver resilient software capability at the speed of relevance. Resilience implies software that is high-quality to produce a portfolio of software capabilities enabled by DOD processes.
Goals (1) innovate for competitive advantage; (2) optimize for efficiencies and improved capability; (3) evolve cybersecurity for an agile and resilient defense posture; and (4) cultivate talent for a ready digital workforce.	**Goals** (1) incorporate engineering and software development earlier in the acquisition life cycle; (2) adopt an integrated framework of shared resources; (3) transform the software workforce; and (4) align software science and technology with acquisition.	**Goals** (1) accelerate the DOD enterprise cloud environment; (2) establish a department-wide software factory ecosystem; and (3) transform processes to enable resilience and speed.
Intent Provide a roadmap to support the implementation of National Defense Strategy priorities.	**Intent** Address statutory requirements to, among other things, outline a plan to advance and enable the rapid transition of software-developed capabilities to acquisition programs through research and development and science and technology initiatives.	**Intent** Establish a path to deliver resilient software capability quickly. This strategy addresses aspects of the Digital Modernization Strategy.

Source: GAO analysis of DOD documentation | GAO-23-105611

Figure 5: Visions, Goals, and Intent of the Department of Defense's (DOD) Key Software Modernization Strategies

The plans further define each goal through objectives or focus areas. For example:

- To achieve its goal of establishing a department-wide software factory ecosystem, DOD outlines five key objectives in its Software Modernization Strategy, such as advancing DevSecOps through enterprise providers and accelerating software deployment with continuous authorization.
- To achieve its goal of transforming the software workforce, DOD outlines five focus areas in its Software Science and Technology Strategy—training and investing in data science, artificial intelligence, machine learning, and software engineering as well as cultivating a software engineering workforce.

According to DOD, these future software modernization efforts are expected to require sustained effort to fully implement. For example, DOD's Software Modernization Strategy states that software modernization is a continuous journey where success requires action and a shift in mindset and culture. In addition, Office of the USD(A&S) and DOT&E officials said that it will take time to develop and encourage the adoption of Agile software practices across the department and establish supporting infrastructure, such as training the software development, acquisition, and cybersecurity workforce in modern software methods.

DOD Has Yet to Fully Implement Key Practices to Facilitate Future Software Modernization Plans

In its preparation to implement future software modernization efforts, DOD fully or substantially followed two of six, partially followed three, and has yet to implement one of six selected practices that our prior work shows can help agencies implement transformative changes. While DOD incorporated some elements of these four practices, we found gaps in the implementation of each.

DOD has substantially followed key practices related to involving employees and key stakeholders, and employee engagement.

Involving employees and key stakeholders. DOD took steps or developed plans to involve Congress, key stakeholders, such as the private sector, and employees in developing software modernization reforms. Our prior work shows that involving employees and key stakeholders helps facilitate goals, incorporate insights, and increase acceptance of transformation change. Examples of DOD's related efforts include:

- Software acquisition pathway. DOD has continuously involved employees in developing and refining aspects of the software acquisition pathway. OSD established a working group that collaborates with the military departments and other DOD organizations to shape policies and guidance related to the implementation of the pathway, according to officials from the Office of the USD(A&S). Additionally, the Office of the USD(A&S) continues to iteratively deploy guidance to aid programs transitioning to the pathway, including regularly updating policy and guidance and resources for the software acquisition pathway on DOD's Adaptive Acquisition Framework website. Officials from the Office of the USD(A&S) noted that these resources incorporate lessons learned and are intended to aid the software workforce in effectively delivering and acquiring software through the pathway. They added that they also consult directly with programs considering the pathway and plan to continue to do so as the pathway evolves.
- Software and Digital Technology Pilot Program. In December 2020, the Consolidated Appropriations Act of 2021 established the Software and Digital Technology Pilot program. The Office of the USD(A&S), in collaboration with the Office of the USD(C), engaged Congress to help establish the pilot program. Office of the USD(C) officials told us they proposed the single appropriation category to Congress after receiving initial support from within DOD. They noted that they continue to engage with Congress regarding proposals to expand the pilot, which began in fiscal year 2021. However, Congress has yet to approve any additional programs to date. DOD intends to execute the pilot for several years and subsequently work with Congress to implement a long-term funding solution.
- Ignite initiatives. According to officials from the Office of the USD(A&S), they established initiatives—which DOD refers to as ignite initiatives—with a goal of transforming functions such as requirements, cost estimating, and test and evaluation processes for software. The officials said that these initiatives include representatives from Joint Staff, OSD, and the military departments to provide input on policies, processes, and culture to enable modern software delivery.

DOD has also involved industry stakeholders in developing reforms. For example, DOD collaborated with industry to develop the Continuous Iterative Development Measurement Framework, which is a comprehensive set of metrics to evaluate vendor software factories.

DOD also has plans to involve additional stakeholders in future reforms, such as by partnering with industry to improve contracting processes and ensure access to enterprise cloud services. Two of DOD's key strategies establish goals and objectives related to working with industry, such as on cloud capabilities. For example, the Digital Modernization Strategy states that DOD will partner with industry to securely deliver cloud capabilities in alignment with mission requirements to achieve its goals. Further, the Software Modernization Strategy notes that DOD must partner with industry to improve contracting processes for cloud services, including a range of enterprise contracts that leverages existing acquisition success while avoiding duplication.

Employee engagement. DOD has taken several actions to sustain and strengthen employee engagement for its future software modernization reforms, such as educating employees, conducting targeted outreach, and forming working groups. Our past work emphasizes the importance of this step because people define the organization's culture and drive its performance. Examples of DOD's efforts to engage employees include:

- DOD has communicated with employees on software modernization reform efforts. For example, the Office of the USD(A&S) performed outreach to and developed guidance for individual program offices to facilitate their transition to modern software approaches. OSD offices also offered training, such as through conferences and webinars, to educate the workforce on modern software approaches and why and how DOD needs to fundamentally transform the way it develops and acquires software.
- DOD and the military departments encourage participation in software communities of practice to share best practices and lessons learned on modern software approaches.
- According to an official from the Office of the USD(R&E), the office continuously engages with software factory stakeholders, such as the Office of the USD(A&S), DOD CIO, and software acquisition programs, at formal presentations and forums to understand what support software factories need from OSD organizations. These discussions include working with programs to help eliminate barriers for software factories.
- The Software Modernization SSG established an Action Officer Working Group that includes representatives from across DOD organizations and the military departments to help coordinate future software modernization initiatives.

DOD Developed Outcome- Oriented Goals but Has Yet to Establish Performance Measures

DOD has partially followed a key practice related to establishing goals and outcomes. Our past work has found that agencies should establish clear outcome-oriented goals to help identify what they are trying to achieve with their reform efforts and should establish performance measures to assess the extent to which they are meeting their goals. DOD's key department-wide strategies for software modernization establish clear outcome-oriented goals and objectives that align with DOD's mission and strategic plans, such as the National Defense Strategy. For example:

- DOD's Digital Modernization Strategy outlines a goal to preserve and expand the U.S. military's competitive advantage against adversaries. This goal depends on the United States' ability to deliver technology faster, a theme throughout the 2018 National Defense Strategy. Specifically, the National Defense Strategy notes that continuously delivering performance with affordability and speed is a defense objective.
- DOD's Software Modernization and Science and Technology strategies state that software modernization requires the department to transform its software workforce to adopt the appropriate technical skills, such as equipping software engineers, developers, and testers with modern tool sets, processes, and capabilities. These efforts align with cultivating workforce talent, as discussed in the 2018 National Defense Strategy. Specifically, the National Defense Strategy notes that cultivating a lethal force relies on the ability of warfighters and others in DOD's workforce to integrate new capabilities, adapt warfighting approaches, and change business practices to achieve mission success. The Software Modernization Strategy states that DOD's workforce must understand its role in delivering software, streamline processes, push for automation, and better leverage technology.

However, DOD has yet to establish performance measures to assess progress toward its goals. According to DOD officials, the department is developing implementation plans that are expected to include performance measures. Specifically, officials told us the Software Modernization Strategy implementation plan will include performance measures to assess progress against priority tasks, which track to outcome-oriented goals. DOD officials noted in November 2022 that the Software Modernization Strategy implementation plan is in draft and is expected to be published in the second quarter of fiscal year 2023. The Software Science and Technology Strategy states that its implementation plan will, among other things, establish and define metrics for outcome- oriented goals. According to DOD officials, the Software Science and Technology Strategy implementation plan is being drafted, with an estimated publication date in the first or second quarter of calendar year 2023.

While its plans to include performance measures in implementation plans are a positive step, DOD has yet to identify the steps it will take to develop effective measures. We have previously identified key attributes of successful performance measures, such as linkage to an agency's goals, which help organizations track the progress they are making and assess whether performance is meeting expectations (see appendix VI). DOD's key strategies do not establish any guidelines for the characteristics of performance measures to be developed. DOD officials noted they had yet to determine the particular measures they would use to assess progress against outcome-oriented goals because the plan is still in draft. As DOD finalizes implementation plans for its future software modernization efforts, ensuring that key attributes of successful performance measures are included, as appropriate, will help guarantee that DOD is well positioned to assess progress against outcome-oriented goals. In turn, the ability to assess progress will help DOD course correct, if necessary, to reach the desired software modernization outcomes.

DOD Established an Implementation Team but Has Yet to Fully Identify Resources or Responsibilities

DOD has partially followed a key practice related to leadership focus and attention. Our prior work shows that providing leadership for transformational reforms includes several things, such as establishing a dedicated implementation team with sufficient resources, designating leaders responsible for implementation, and holding those leaders accountable. DOD has established an implementation team but has yet to identify the resources needed to lead DOD's software modernization efforts or fully determine how it will hold department leaders engaged in these efforts accountable.

Dedicated implementation team with capacity to manage reforms. DOD has established a dedicated implementation team to manage its software modernization reform process. The Software Modernization SSG is the main governance body that oversees and leads the implementation of software modernization reforms across DOD, including activities supporting the Software Modernization Strategy.

While DOD officials told us that individual working groups are assessing the requirements to execute key areas of the Software Modernization Strategy, DOD has yet to take steps to determine whether the Software Modernization SSG as a whole will have the capacity and resources necessary to lead software modernization activities. The Software Modernization SSG relies on its members from OSD organizations, the Joint Staff, and the military departments to identify the resources each member organization is able to devote to support software modernization implementation. DOD officials noted that these entities must balance their own ongoing organizational commitments with available staffing and resources to support software reform efforts.

Identifying needed staffing and resources for DOD's dedicated implementation team could help DOD ensure that the Software Modernization SSG can effectively carry out its leadership role in implementing software modernization efforts.

Assigning leadership roles and responsibilities and holding leaders accountable. DOD's current planning documentation broadly assigns high-level leadership responsibility for implementing software modernization reforms. For example, DOD's Software Modernization SSG is tri-chaired by senior representatives from Offices of the USD(A&S), USD(R&E), and DOD CIO. These organizations are tasked with leading collaboration with other DOD organizations and the military departments as well as making decisions related to DOD's software modernization activities. Additional membership of the Software Modernization SSG includes representatives from across DOD, including DOT&E, CAPE, Joint Staff, and the military departments. These organizations and departments are to provide representation in all efforts pertaining to modern software development and delivery.

DOD's Software Modernization Strategy states that software modernization requires a cohesive departmental effort that involves various DOD organizations. The strategy states that implementation success depends heavily on partnerships and collaboration across the department given the role and pervasiveness of software across mission capabilities and supporting infrastructure. Further, the Deputy Secretary of Defense's February 2022 memorandum approving the strategy stated that all offices and personnel are expected to provide the necessary support for software modernization.

However, DOD has yet to fully develop an approach to hold accountable the many leaders who will need to be involved in implementing software modernization reforms. This is in part because DOD has yet to fully identify in key documents what entities will be involved in executing software modernization efforts and what their specific responsibilities will entail. For example, DOD's current planning documentation, including the Software Modernization and Software Science and Technology strategies, do not address the specific responsibilities of OSD offices with leadership roles or of the military departments and other organizations involved in implementation.

According to DOD officials, once issued, the Software Modernization Strategy implementation plan will identify an Office of Primary Responsibility to support key lines of effort. For example, individual DOD organizations and military departments will be responsible for implementing modern software practices, such as cloud computing and DevSecOps, at the program- and component-levels. The Software Modernization SSG is expected to monitor the efforts of these organizations. Office of the USD(A&S) officials noted that software modernization at DOD relies heavily on the DOD organizations and military departments.

While assigning lead offices is an important step in implementation planning, this approach, as described by DOD, does not ensure that DOD will fully identify the specific roles and responsibilities of leaders involved in transformational software reforms. Until DOD fully identifies the roles and responsibilities for these leaders, DOD will likely be challenged to hold them accountable for implementation.

DOD Is Developing Implementation Plans but Has Yet to Identify Data Collection Methods for Monitoring Progress

DOD has yet to implement a key practice related to managing and monitoring implementation. Our prior work emphasizes the importance of developing an implementation plan with key milestones and deliverables and putting in place processes to collect the needed data and evidence to effectively measure the reforms' outcome-oriented goals. DOD is in the process of developing implementation plans for its key strategies, although these plans have been delayed from their original planned release dates. Further, DOD has yet to describe how the department plans to collect the data necessary to measure progress in achieving strategic goals.

Developing implementation plans. According to DOD officials, the implementation plans they are developing for the Software Modernization and Software Science and Technology strategies are expected to include key milestones and deliverables to track implementation progress. For example, DOD officials told us that the Software Modernization Strategy implementation plan will include a governance structure to assess, reprioritize, and track progress toward goals, such as measurable deliverables and milestones per activity outlined in strategic goals.

However, DOD has yet to publish these plans and has already delayed its anticipated completion dates for the Software Modernization Strategy. The February 2022 approval memorandum from the Deputy Secretary of Defense for the Software Modernization Strategy directed the delivery of an implementation plan within 180 days, which would have been in August 2022. The planned completion date for this plan has now slipped to the second quarter of fiscal year 2023. According to a DOD official, delays in publishing the implementation plan are due to the need for additional time for internal coordination among DOD leadership to clear for publication. Further, DOD officials told us that the Software Science and Technology Strategy implementation plan is expected to be published after the Software Modernization Strategy implementation plan to, in part, ensure that the goals outlined in both plans align. Given the importance of these plans in helping to manage and monitor implementation, it is essential that DOD finalizes them in a timely manner.

Processes and data to measure effectiveness of reforms. DOD has yet to describe how the department plans to collect the data necessary to effectively assess its progress against performance measures. According to DOD officials, the department plans to collect data to measure performance and expects to analyze it in Advana—DOD's enterprise data platform. However, DOD officials have yet to fully identify the methods they plan to use to collect data across the department or specify how they plan to use the data collected, in part, because DOD's data collection efforts related to software modernization to date have focused on the software acquisition pathway.

DOD Instruction 5000.87, Operation of the Software Acquisition Pathway, requires pathway programs to report data to assess and manage program performance and progress, such as average lead time and value assessment rating. However, software acquisition pathway program metrics and reporting requirements apply to a selected group of programs out of many in the department that are developing or acquiring software. Further, the software acquisition pathway is one component of DOD's software modernization efforts outlined in department-wide software strategies and does not represent the breadth of planned software modernization efforts.

Developing implementation plans for the Software Modernization and the Software Science and Technology strategies and establishing processes to collect the necessary data and evidence will help DOD ensure it is well positioned to measure progress toward implementing its goals.

DOD Has Yet to Conduct Strategic Planning for Its Software Workforce

DOD has partially followed a key practice related to strategic workforce planning. Our prior work has found that agencies should complete this planning to ensure they have the needed resources and capacity to successfully execute reforms. DOD has taken initial steps to identify its software workforce, a crucial effort that must be completed prior to conducting strategic workforce planning. However, it has yet to determine whether it has the needed workforce resources and capacity to successfully execute planned software modernization reforms.

According to DOD, a workforce skilled in modern software development practices is fundamental to carrying out software modernization efforts. DOD's Software Modernization Strategy states that modern software practices require a shift in DOD's workforce and that developing, training, and recruiting that workforce are critical elements of software modernization. Both DOD's Software Science and Technology and Digital Modernization strategies identify transforming DOD's software workforce as a key goal.

Identifying the software workforce. DOD is taking initial steps to identify the makeup of its current software workforce. According to officials from the Office of the USD(A&S), determining the composition of the software workforce, such as identifying DOD professionals that currently make up the software workforce and the additional roles that would be needed to successfully adopt department-wide reforms, has been a challenge. A 2020 RAND study noted that DOD lacks a workforce model that properly supports a software acquisition workforce, such as an official software career field or a system for identifying or tracking software professionals in the department. This study included a recommendation for the department to identify who is in the software acquisition workforce and presented options for DOD to track and manage this workforce.

In July 2021, the department established the Digital Talent Management Forum, which aims to identify and define key software engineering roles needed for modern software delivery, according to DOD officials. These officials noted that the forum is supporting DOD CIO's efforts to expand the DOD Cyber Workforce Framework to include software engineering and software testing roles in the framework's database.

An official from the Office of the Under Secretary of Defense for Personnel and Readiness explained that, through this effort, the department is working to collect data to identify software professionals across DOD's workforce, such as those performing software functions that may not be captured in a job title or occupational series. The official noted that identifying the software workforce is currently a challenge for DOD because software professionals work across many occupational series. Once DOD captures the data, officials expect it will provide department-wide information on the software workforce composition, expertise, and skill sets. DOD officials said this data capture effort is expected to take about 12 to 18 months. The resulting insight into the composition of its software workforce should help DOD determine what resources are needed to support software modernization reforms.

Conducting strategic workforce planning. While identifying the workforce is a critical step, it is only the first step in a longer process to ensure that DOD will have the workforce it needs to execute its software modernization reforms. Key principles for strategic workforce planning in our prior work state that this planning should address two critical needs:

1) aligning an organization's human capital program with its current and emerging mission and programmatic goals and 2) developing long-term strategies for acquiring, developing, and retaining staff to achieve programmatic goals. Figure 6 illustrates the strategic workforce planning process.

DOD has yet to determine how it will execute this broader strategic workforce planning process for its software modernization efforts. DOD officials acknowledged that data collection is only the first step in conducting workforce planning. They noted that once software workforce professionals are properly identified in personnel data, DOD can conduct a workforce capability assessment. However, officials noted that DOD is still in the early stages of these identification efforts. Similarly, an Office of the Under Secretary of Defense for Personnel and Readiness official noted that DOD is currently focused on elements that must be in place before strategic workforce planning can begin, such as determining the critical skills and competencies the software acquisition workforce needs to achieve programmatic results.

Strategic workforce planning for software modernization efforts is likely to take a number of years and will need to involve the coordinated efforts of management, employees, and key stakeholders across DOD. Developing a department-wide strategic workforce plan for DOD's software workforce—including strategies tailored to address gaps in the critical skills and competencies—will help position DOD to execute next steps in this planning process and achieve future software modernization goals.

Conclusions

DOD has made numerous efforts to modernize its software acquisition and development approaches in recent years, but much work remains in this crucial area. DOD's recently issued software strategies include ambitious goals that are essential to moving from early adoption of modern software practices by selected programs to a lasting, department- wide transformation. Meeting these goals will improve DOD's ability to keep pace with strategic competitors, such as Russia and China.

As DOD begins to translate its goals into action, incorporating key change management practices identified in our past work will help senior leadership oversee continued progress towards software transformation. For example, taking action to develop meaningful performance measures, establish data collection strategies for measuring performance, and finalize implementation plans can help DOD track progress towards achieving and implementing software modernization goals. Moreover, establishing a sufficiently-resourced implementation team and delineating roles and responsibilities associated with software modernization efforts can help ensure that leaders have the resources they need to implement reforms and are held accountable for achieving them.

Further, building a workforce—with critical skills and competencies—that can implement these reforms is foundational to all of DOD's planned actions. Until DOD determines when and how it will conduct effective workforce planning for its software workforce, its ability to implement its planned actions and meaningfully transform its software acquisition practices as intended remains in question.

Recommendations for Executive Action

We are making the following seven recommendations to DOD:

The Secretary of Defense should ensure that, as the Software Modernization SSG and other relevant entities develop performance measures for future software modernization efforts, these measures incorporate GAO's key attributes of successful performance measures, to the extent appropriate, to track progress towards achieving agency goals. (Recommendation 1)

The Secretary of Defense should direct the USD(A&S), USD(R&E), and DOD CIO to identify the resources needed, such as staffing and funding, to lead DOD's software acquisition and development reform efforts, and to address any related deficiencies these officials identify. (Recommendation 2)

The Secretary of Defense should fully identify roles and responsibilities for leaders throughout the department for carrying out reforms included in key software strategies. (Recommendation 3)

The Secretary of Defense should ensure the USD(A&S), USD(R&E), and DOD CIO finalize an implementation plan that includes key milestones and deliverables to track progress on implementing the Software Modernization Strategy. (Recommendation 4)

The Secretary of Defense should ensure the USD(R&E) finalizes an implementation plan that includes key milestones and deliverables to track progress on implementing the Software Science and Technology Strategy. (Recommendation 5)

The Secretary of Defense should direct the USD(A&S), USD(R&E), and DOD CIO to establish processes to collect the data necessary to effectively measure progress against outcome-oriented goals related to software modernization efforts. (Recommendation 6)

The Secretary of Defense should ensure that, once the software workforce is identified, the USD(A&S), the Under Secretary of Defense for Personnel and Readiness, and other relevant entities, use that information to develop a department-wide strategic workforce plan that identifies strategies tailored to address gaps in the critical skills and competencies needed to achieve software modernization goals. (Recommendation 7)

PANEL 22. OPPORTUNITIES WITH MODELING & SIMULATION

Thursday, May 9, 2024	
3:45 p.m. – 5:00 p.m.	**Chair: Lieutenant Colonel Charles P. Rowan, USA,** Assistant Professor, Interim Director of MOVES Institute, Naval Postgraduate School ***Statistical Procedures for Validation of a Computer Model with Multimodal Output When the Observation is a Single Time Series*** Patricia Jacobs, Naval Postgraduate School ***Challenges and Opportunities in Enhancing Department of Defense Ground Vehicle Capabilities through Digital Transformation*** Waterloo Tsutsui, Purdue University ***A Model for Evaluating the Maturity of a Modular Open Systems Approach*** Alfred Schenker, Carnegie Mellon Univrsity

Lieutenant Colonel Charles P. Rowan, USA—was commissioned as an armor officer from the United States Military Academy at West Point in 2005. LTC Rowan first served in a variety of leadership and staff positions with the 3rd Infantry Division at Fort Stewart, Georgia, including a deployment to Operation Iraq Freedom 07-09 as a tank company executive officer in Baghdad, Iraq. LTC Rowan was next assigned to the 1st Armored Division at Fort Bliss, Texas. He served as a battalion staff officer and then as an infantry company commander while deployed to Operation Iraqi Freedom 09-10 in Kirkuk, Iraq.

LTC Rowan was selected to be an instructor and later Assistant Professor in the Department of Behavioral Sciences and Leadership at West Point. LTC Rowan then transferred to Functional Area 57, Simulation Operations. He was assigned as the Division Simulation Operations Officer for the 2nd Infantry Division at Camp Red Cloud, Republic of Korea. He then served as the Chief of Knowledge Management for the 25th Infantry Division and US Army Hawaii at Schofield Barracks, Hawaii. LTC Rowan currently serves as an Assistant Professor and Interim Director of the Modeling, Virtual Environments, and Simulation (MOVES) Institute at the Naval Postgraduate School.

LTC Rowan holds a Ph.D. in Modeling, Virtual Environments, and Simulation and a Master of Science degree in Human Systems Integration, both from the Naval Postgraduate School, and a Bachelor of Science degree in Engineering Psychology from the United States Military Academy.

LTC Rowan's awards and decorations include the Bronze Star Medal with Oak Leaf Cluster, the Meritorious Service Medal with three Oak Leaf Clusters, the Army Commendation Medal, the Army Achievement Medal with four Oak Leaf Clusters, the Combat Action Badge, the Air Assault Badge, and the Order of St. George Bronze Medallion.

Statistical Procedures for Validation of a Computer Model with Multimodal Output When the Observation is a Single Time Series

Patricia (P.A.) Jacobs—is a retired Distinguished Professor of Operations Research at the Naval Postgraduate School. She is a fellow of the Royal Statistical Society (RSS), the American Statistical Association (ASA), and the American Association for the Advancement of Science (AAAS). [pajacobs@nps.edu]

Abstract

The AEGIS combat system is a system of systems (SoS) in which the systems and tactics for their use continually evolve. The Combat System Test Bed (CSTB) is a federation of computer simulation models (SMs) that represents the performance of the AEGIS SoS. Physical test events are conducted to assess how well the CSTB represents the performance of the evolving AEGIS SoS. Test event data, including time series, are compared to SMs' output. In this paper, simulation is used to study the efficacy of statistical procedures, including one currently in use, to obtain statistical evidence that a model's multimodal distribution does not include that of a time series observation.

Introduction

The AEGIS combat system is a system of systems (SoS) in which the systems and tactics for their use continually evolve. The Combat System Test Bed (CSTB) is a federation of computer simulation models (SMs) that represent the performance of the AEGIS SoS. Physical test events are conducted to assess how well the CSTB represents the performance of the evolving AEGIS SoS. Measurements from the test events are compared to CSTB output. Statistical evidence that the CSTB does not well summarize the test event's measurements can lead to CSTB modification or enhancement.

Su et al. (2022) present results comparing four statistical procedures, including one currently used, for computer model validation in the case that the test event measurement is a single time series. In this working paper, we consider additional statistical procedures to validate a computer model in the case the test event measurement is a single time series and the model output is multimodal. The validation of a computer model for test events resulting in a time series has been of interest in several areas, including meteorology (cf. Gneiting et al., 2008) and economics (cf. Diebold et al., 1998, 1999).

The next section, Procedures to Assess How Well a Multimodal Model Distribution Summarizes One Observed Time Series, presents three procedures to assess how well a computer model with multimodal output summarizes one observation time series. The procedures summarize the model replications at each time and compare the observation time series to the summaries to obtain statistical evidence that the model's multimodal distribution does not include that of the observation. One procedure considered is the currently used two-sided hypothesis procedure described in Su et al. (2022) that uses the sample mean and sample variance of the model replications at each time to create 2-sided confidence intervals; we call this procedure the 2.5-sigma procedure. The second procedure uses a 2-sided 99% percentile confidence interval of the model replications at each time. The third type of procedure uses a kernel density estimate (KDE) of the model's density function at each time and the value of the model's KDE of the observation at that time. The third section, Simulation Study, presents the results of a simulation study of the efficacy of the procedures when the model has a multimodal mixture distribution.

The simulation results suggest that the 2.5-sigma procedure is among the least likely to result in a false positive; a false positive occurs when the procedure results in statistical evidence that the multimodal model mixture distribution does not include the distribution of the observation when the model's distribution includes that of the observation. However, the 2.5-sigma procedure is among the least likely of the procedures considered to result in correct statistical evidence that the model's distribution does not include that of the observation. The percentile confidence procedure is the most likely to result in a false positive; it is more likely than the 2.5-sigma procedure to result in correct statistical evidence that multimodal model distribution does not include that of the observation. The KDE procedure is better at balancing incorrect and correct statistical evidence that the model distribution does not include that of the observation.

Procedures to Assess How Well a Multimodal Model Distribution Summarizes One Observed Time Series

Let $X(t_k)$ be the value of the test event's observed time series at time $t_k, k = 1,..., n_T$. Let $Y_i(t_k)$ be the value of the *ith* model replication at time t_k for $i = 1,..., n_M$ where n_M is the number of model replications.

The 2.5-Sigma Procedure and Percentile Interval Procedure

The two-sided hypothesis procedure of Su et al. (2022) is as follows: for each time t_k, the sample mean and sample standard deviation of the n_M model replications $\{Y_i(t_k); i = 1,...., n_M\}$ are computed. The 2.5-sigma confidence interval has lower bound equal to the sample mean minus 2.5 times the sample standard deviation and upper bound equal to the sample mean plus 2.5 times the sample standard deviation. There is said to be statistical evidence that the model's multimodal distribution does not contain that of the observation if the fraction of times the observed time series is outside of the confidence intervals is greater than 0.1. We call this two-sided hypothesis procedure the 2.5-sigma procedure. If $\{Y_i(t_k); i = 1,..., n_M\}$ are independent and identically distributed having a Gaussian distribution, then the 2.5-sigma confidence interval is an approximate 99% prediction interval. This procedure is an approved simulation validation method at the Johns Hopkins University Applied Physics Laboratory (APL; cf. Su et al., 2022).

For each time t_k, let $q_L(t_k)$ (respectively $q_H(t_k)$) be the 0.005 quantile (respectively 0.995 quantile) of the n_M model replications at time t_k, $\{Y_i(t_k); i = 1,..., n_M\}$. The 99% percentile interval is $[q_L(t_k), q_H(t_k)]$. There is said to be statistical evidence that the model's multimodal distribution does not contain that of the observation if the fraction of times the observed time series is outside of the confidence intervals is greater than 0.1.

Gaussian Kernel Density Estimation (KDE) Procedure

- ***Introduction***

Let $F_{Y(t)}(y) = P\{Y(t) \le y\} = \int_{-\infty}^{y} f_{Y(t)}(z)\,dz \text{ for } -\infty < y < \infty$ be the model's probability cumulative distribution function for time t; $f_{Y(t)}(\bullet)$ is the probability density function of the model at time *t*. A Gaussian kernel density smoothing of the model replications at time t,

$\kappa_{Y(t)}(\bullet)$, is an estimate of $f_{Y(t)}(\bullet)$ (cf. Silverman, 1986). Let $\kappa_{Y(t)}(X(t))$ be the value of the model's kernel density estimate (KDE) evaluated at the value of the observed time series at time t. For small positive *h*, $\kappa_{Y(t)}(X(t))h$ is an estimate of $F_{Y(t)}(X(t)+(h/2))-F_{Y(t)}(X(t)-(h/2))$; the model's probability that the value of the observed time series at time t occurs in the interval $[X(t)-(h/2),X(t)+(h/2)]$. There is statistical evidence that the distribution of the observed time series is not included in that of the model if the model's KDE for the observation at time t_k, $\kappa_{Y(t_k)}(X(t_k))$, is too small for too many times $t_k, k=1,...,n_T$.

- ***A Kernel Density Estimate (KDE) Procedure***

For the *ith* model replication at time t_k, $Y_i(t_k)$, let $\kappa_M(\cdot;i,t_k)$ be the Gaussian kernel density estimate (KDE) obtained using model replications at time t_k **without** the ith model replication, $\{Y_j(t_k); j\in\{1,...,i-1,i+1,...,n_M\}\}$. Let $\kappa_M(Y_i(t_k);i,t_k)$ be the value of the KDE for time t_k at $Y_i(t_k)$. Let $q_M(t_k;\alpha)$ be the $\alpha-$quantile of $\{\kappa_M(Y_i(t_k);i,t_k);i=1,...,n_M\}$, the KDEs of the n_M model replication values at time t_k. Let $Q_M(i;\alpha)=\sum_{k=1}^{n_T} I(\kappa_M(Y_i(t_k);i,t_k)<q_M(t_k;\alpha))$, the number of times the value of the KDEs for the *ith* model replication are less than the model values' KDE $\alpha-$quantiles; $I(A)$ equals 1 if event *A* occurs and 0 otherwise.

For each time t_k, let $\kappa_A(\cdot;t_k)$ be the KDE obtained using all the model replications at time t_k, $\{Y_i(t_k);i=1,...,n_M\}$. Let $\kappa_O(t_k)=\kappa_A(X(t_k);t_k)$, the value of the KDE, $\kappa_A(\cdot;t_k)$, at $X(t_k)$, the value of the observed time series at time t_k. Let $Q_O(\alpha)=\sum_{k=1}^{n_T} I(\kappa_O(t_k)<q_M(t_k;\alpha))$, the number of times the observed time series KDEs are less than the model values' KDE $\alpha-$quantiles. $Q_O(\alpha)$ is compared to a lower bound obtained from the model replications. The most conservative lower bound considered is $B_M(\alpha;1)=\max_{i=1,...,n_M}(Q_M(i;\alpha))$, the maximum number of times a model's replication value KDEs are less than the model values' KDE $\alpha-$quantiles. Other considered lower bounds are $B_M(\alpha;n)=$ the nth largest number of times a model's replication value KDEs are less than the model values' KDE $\alpha-$quantiles for *n*=2, 3, 4.

For a chosen lower bound, there is statistical evidence that the multimodal distribution of the model does not include that of the observation if the number of times the observed time series KDEs are less than the model values' KDE $\alpha-$quantiles, $Q_O(\alpha)$, is greater than the chosen lower bound.

Simulation Study

There are 500 simulation replications. Each simulation replication generates 300 model replications. The model replications have a mixture distribution. Nine observation time series each having different parameter values are also generated. All the considered model validation procedures compare each of the 9 observed time series to the same model replications.

Simulated Model and Observation

The ith replication of the simulation model satisfies

$$\boldsymbol{Y}_i\left(t_k\right)=1-\exp\left\{-\boldsymbol{\theta}_i\left(M\right)\times\frac{k}{100}\right\}+\boldsymbol{\varepsilon}_i\left(k;M\right)\ \text{for}\ k=1,\ldots,100 \quad (1)$$

where $\left\{\boldsymbol{\theta}_i\left(M\right);i=1,\ldots,n_M\right\}$ are independent identically distributed random variables having a mixture distribution; with probability 0.5, $\boldsymbol{\theta}_i\left(M\right)$ has a gamma distribution with shape parameter $\beta_i\left(M\right)=50$ and mean $\beta_i\left(M\right)/\alpha_i\left(M\right)=2$; with probability 0.5, $\boldsymbol{\theta}_i\left(M\right)$ has a gamma distribution with shape parameter $\beta_i\left(M\right)=50$ and mean $\beta_i\left(M\right)/\alpha_i\left(M\right)=6$; $\left\{\boldsymbol{\varepsilon}_i\left(k;M\right);k=1,\ldots,100\right\}$ are independent identically distributed random variables having a normal distribution with mean $\mu_M=0$ and standard deviation $\sigma_M=0.02$.

The observed time series satisfies

$$\boldsymbol{X}\left(t_k\right)=1-\exp\left\{-\boldsymbol{\theta}\left(O\right)\times\frac{k}{100}\right\}+\boldsymbol{\varepsilon}\left(k;O\right)\ \text{for}\ k=1,\ldots,100 \quad (2)$$

where $\boldsymbol{\theta}\left(O\right)$ is a random variable having a gamma distribution with shape parameter $\beta\left(O\right)=50$ and mean $\frac{\beta\left(O\right)}{\alpha\left(O\right)}$; $\left\{\boldsymbol{\varepsilon}\left(k;O\right);k=1,\ldots,100\right\}$ are independent identically distributed random variables having a normal distribution with mean $\mu_O=0$ and variance σ_O^2. The values of the parameters, $\left(\frac{\beta\left(O\right)}{\alpha\left(O\right)},\sigma_O\right)$, considered are (1, 0.02), **(2, 0.02)**, (2, 0.1), (3, 0.02), (4, 0.02), (5, 0.02), **(6, 0.02)**, (6, 0.1) and (10, 0.02); the bold values correspond to parameters included in the model mixture distribution.

Kernel Density Estimation

The software R function called density with Gaussian kernel and bandwidth ucv (unbiased cross validation) is used to obtain the kernel density estimates (cf. R Core Team, 2021).

Figure 1 displays fifty model replications as a function of time. Figure 2 displays a histogram of 300 model replications at time 0.5. Figure 3 displays the kernel density estimate using the model replications at time 0.5. The three Figures suggest that the model replications have 2 modes. Figure 4 displays the number of times the model replications value KDEs, $\left\{\kappa_M\left(Y_i\left(t_k\right);i,t_k\right);k=1,\ldots,100\right\}$, are less than the model replication values' KDE 0.001 quantiles, $\left\{q_M\left(t_k;0.001\right);k=1,\ldots100\right\}$;

$$Q_M\left(i;0.001\right)=\sum_{k=1}^{n_T}I\left(\kappa_M\left(Y_i\left(t_k\right);i,t_k\right)<q_M\left(t_k;0.001\right)\right)\ \text{for}\ i=1,\ldots,300 \quad (3)$$

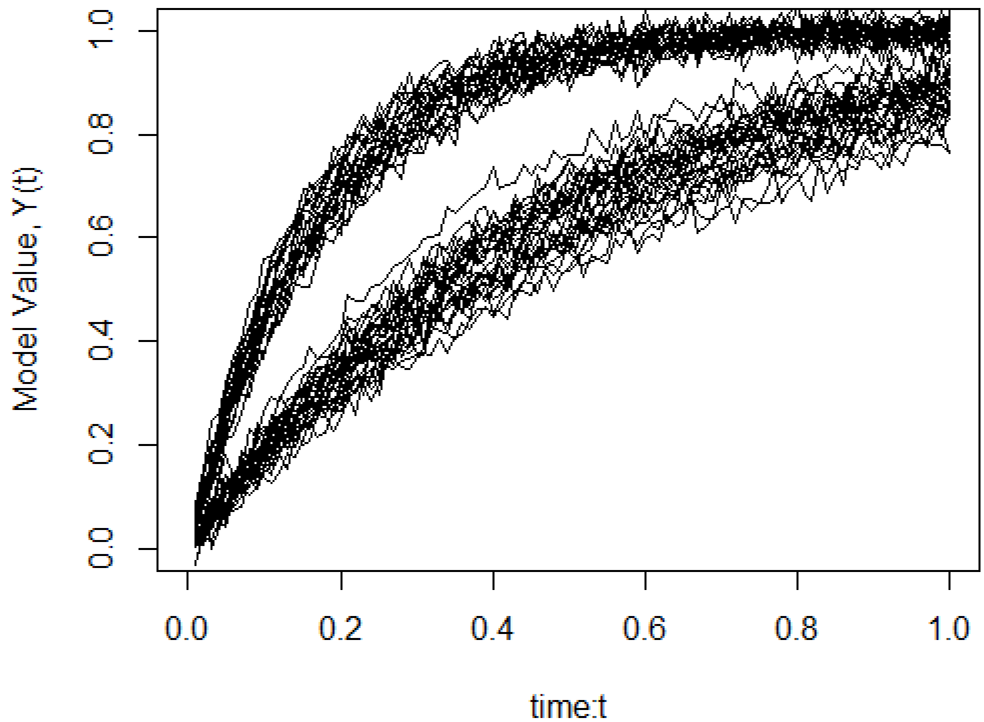


Figure 1. 50 Model Replications

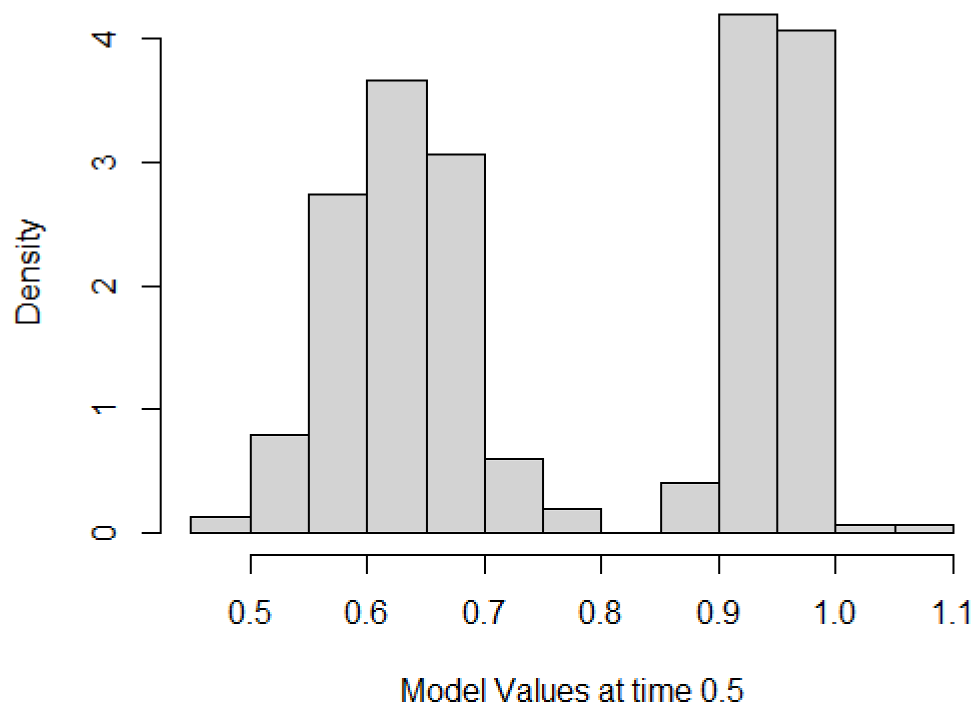


Figure 2. Model Replications at Time 0.5

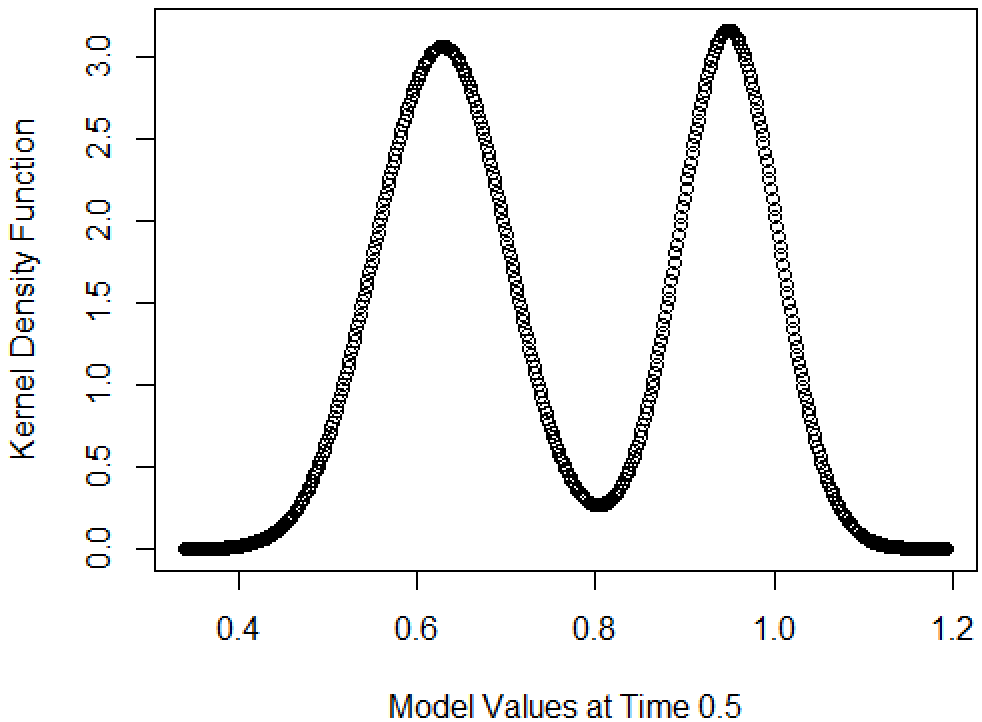


Figure 3. Model Replications at Time 0.5 Gaussian Kernel Density With Bandwidth=ucv (Unbiased Cross Validation)

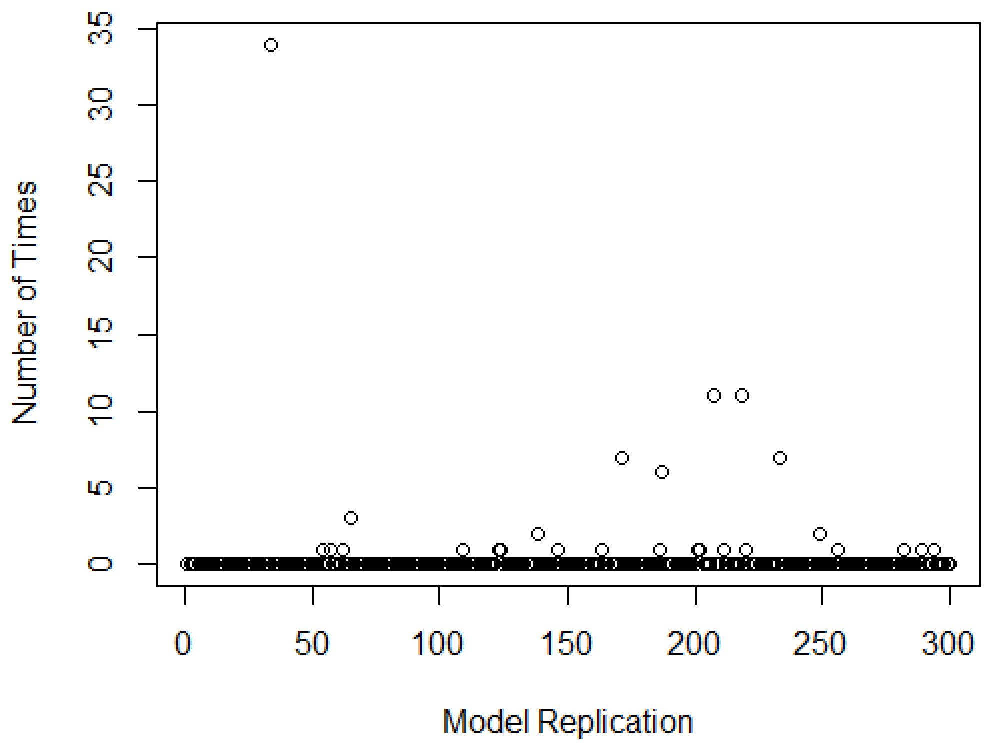


Figure 4. Number of Times Model Replication's Value Kernel Density Estimates (KDEs) Are Less Than the Model Values KDE 0.001 Quantiles

Simulation Results

The 2.5-sigma, 99% percentile confidence interval, and kernel density estimate (KDE) procedures are used to obtain statistical evidence that the multimodal mixture model distribution does not include that of the observation. Several values for the KDE procedure quantile and lower bound are considered; the quantile values are $\alpha \in \{0.01, 0.005, 0.001, 0.0005\}$; the lower bounds are the maximum number of times a model replication value KDEs are less than the model values' KDE $\alpha-$quantile, $B_M(\alpha;1)$; the second largest number of times, $B_M(\alpha;2)$; the third largest number of times, $B_M(\alpha;3)$; and the fourth largest number of times, $B_M(\alpha;4)$. There is statistical evidence that the mixture model distribution does not include that of the observation if the number of times the observation's KDEs are less than the model values' KDE quantiles is greater than the lower bound.

Table 1 displays the fraction of the 1,000 cases that result in a false positive (incorrect statistical evidence that the observation distribution is not included in that of the model when the model distribution includes that of the observation). Each of the 500 simulation replications has 2 cases in which the model distribution includes that of the observation;
$\left(\frac{\alpha(O)}{\beta(O)}, \sigma_O\right) \in \{(2,0.02),(6,0.02)\}$.

Table 1. Fraction of Cases With Incorrect Statistical Evidence That the Multimodal Mixture Model Distribution Does NOT Include That of the Observation When the Model Distribution Does Include That of the Observation

Procedure		Fraction of Cases		Fraction of Cases		Fraction of Cases		Fraction of Cases
2.5-sigma		0.001						
99% Percentile Confidence Interval		0.040						
KDE								
	KDE Quantile α	Observation Number Greater Than Maximum Model Number, $B_M(\alpha;1)$		Observation Number Greater Than 2nd Largest Model Number, $B_M(\alpha;2)$		Observation Number Greater Than 3rd Largest Model Number, $B_M(\alpha;3)$		Observation Number Greater Than 4th Largest Model Number, $B_M(\alpha;4)$
	0.01	0.002		0.006		0.010		0.013
	0.005	0.001		0.003		0.003		0.004
	0.001	0.002		0.005		0.007		0.014
	0.0005	0.002		0.005		0.007		0.010

Table 2 displays the fraction of the 3,500 cases that result in correct statistical evidence that the observation's distribution is not included in that of the model; there are 7 such cases in each simulation replication.

Table 2. Fraction of Cases With Correct Statistical Evidence That the Multimodal Mixture Model Distribution Does NOT Include That of the Observation When the Model Distribution Does NOT Include That of the Observation

Procedure		Fraction of Cases		Fraction of Cases		Fraction of Cases		Fraction of Cases
2.5-sigma		0.173						
99% Percentile Confidence Interval		0.526						
KDE								
	KDE Quantile, α	Observation Number Greater Than Maximum Model Number, $B_M(\alpha;1)$		Observation Number Greater Than 2nd Largest Model Number, $B_M(\alpha;2)$		Observation Number Greater Than 3rd Largest Model Number, $B_M(\alpha;3)$		Observation Number Greater Than 4th Largest Model Number, $B_M(\alpha;4)$
	0.01	0.222		0.368		0.513		0.616
	0.005	0.174		0.300		0.438		0.544
	0.001	0.285		0.519		0.641		0.698
	0.0005	0.269		0.497		0.624		0.681

The results displayed in Tables 1 and 2 suggest the following concerning the ability of the considered statistical procedures to result in correct statistical evidence that the model distribution does not include that of the observation.

- ***False Positive: Incorrect Statistical Evidence That the Mixture Model Distribution Does Not Include That of the Observation When the Model Distribution Does Include That of the Observation***

The 2.5-sigma procedure and the KDE procedure with lower bound the maximum number of times a model replication's value KDEs are less than the model's value KDE 0.005-quantiles, $B_M(0.005;1)$, result in 1 false positive in the 500 simulation replications. The KDE procedure with 0.001- quantile and lower bound the maximum number of times a model replication's value KDEs are less than the model values' KDE quantiles, $B_M(0.001;1)$, results in 2 false positives. The KDE procedure with lower bounds $B_M(0.005;2)$ and $B_M(0.005;3)$ result in 3 false positives. The percentile confidence interval procedure results in the most false positives.

- ***True Positive: Correct Statistical Evidence the Observation Distribution Is Not Included in That of the Model***

The 2.5-sigma procedure and the KDE procedure with lower bound the maximum number of times a model replication's value KDEs are less than the model KDE values' 0.005-quantiles, $B_M(0.005;1)$, result in correct statistical evidence the model distribution does not include that of the observation in less than 18% of the cases. The KDE procedure with lower bound $B_M(0.001,1)$ results in one more false positive than the 2.5-sigma procedure and results in correct statistical evidence that the model distribution does not include that of the observation in slightly over 28% of the cases. The KDE procedure with lower bound $B_M(0.005,3)$ results in 2 more false positives than the 2.5-sigma procedure but results in correct statistical evidence that the model distribution does not include that of the observation in slightly over 40% of the cases. The KDE procedure with lower bound $B_M(0.001;4)$, the fourth largest number of times a model replication value KDEs are less than the model values' KDE 0.001-quantiles, results in the most cases with correct statistical evidence the model distribution does not include that of the observation; however, it also results in the highest number of false positives.

Tables 3 and 4 display the fraction of the 500 simulation replications that result in statistical evidence that the observation distribution is not included in that of the model. Table 3 (respectively 4) displays the results using the 0.005 (respectively 0.001) model values' KDE quantile for the KDE procedure. The bold entries correspond to results in the case that the observation distribution is included in that of the mixture model.

Table 3. Fraction of Simulation Replications Resulting in Statistical Evidence That the Model Mixture Distribution Does Not Include That of the Observation

Model Values' KDE quantile $\alpha = 0.005$ Bold Entries: Observation Distribution is included in the Model Mixture Distribution							
Observation parameters		Procedure					
Gamma mean $\frac{50}{\alpha(O)}$	Normal standard deviation σ_O	2.5-Sigma	Percen-tile	KDE Obser-vation number > model's max $B_M(\alpha;1)$	KDE Obser-vation number > model's 2nd max $B_M(\alpha;2)$	KDE Obser-vation number > model's 3rd max $B_M(\alpha;3)$	KDE Obser-vation number > model's 4th max $B_M(\alpha;4)$
1	0.02	0.974	0.998	0.874	0.942	0.978	0.986
2	**0.02**	**0.002**	**0.042**	**0.002**	**0.006**	**0.006**	**0.008**
2	0.1	0.082	0.736	0.020	0.100	0.232	0.422
3	0.02	0	0	0.130	0.280	0.376	0.472
4	0.02	0	0	0.092	0.206	0.302	0.382
5	0.02	0	0.002	0.010	0.018	0.034	0.050
6	**0.02**	**0**	**0.038**	**0**	**0**	**0**	**0**
6	0.1	0.002	0.988	0.046	0.260	0.608	0.830
10	0.02	0.154	0.958	0.050	0.292	0.534	0.666

Table 4. Fraction of Simulation Replications Resulting in Statistical Evidence That the Model Mixture Distribution Does Not Include That of the Observation

Model Values' KDE quantile $\alpha = 0.001$ Bold Entries: Observation Distribution is included in the Model Mixture Distribution							
Observation parameters		Procedure					
Gamma mean $\frac{50}{\alpha(O)}$	Normal standard deviation σ_O	2.5-Sigma	Percen-tile	Obser-vation number > model's max $B_M(\alpha;1)$	Obser-vation number > model's 2nd max $B_M(\alpha;2)$	Obser-vation number > model's 3rd max $B_M(\alpha;3)$	Obser-vation number > model's 4th max $B_M(\alpha;4)$
1	0.02	0.974	0.998	0.920	0.986	0.990	0.996
2	**0.02**	**0.002**	**0.042**	**0.004**	**0.008**	**0.010**	**0.020**
2	0.1	0.082	0.736	0.106	0.402	0.708	0.878
3	0.02	0	0	0.242	0.450	0.554	0.608
4	0.02	0	0	0.182	0.374	0.464	0.532
5	0.02	0	0.002	0.018	0.048	0.060	0.064
6	**0.02**	**0**	**0.038**	**0**	**0.002**	**0.004**	**0.008**
6	0.1	0.002	0.988	0.264	0.744	0.956	0.994
10	0.02	0.154	0.958	0.266	0.626	0.758	0.814

The results displayed in Tables 3 and 4 suggest that the KDE procedure with lower bound the maximum number of times a model replication's value KDEs are less than that of the model values' KDE 0.001-quantiles, $B_M(0.001;1)$, is the best at balancing incorrect and correct statistical evidence the model distribution does not include that of the observation; it results in one more false positive than the 2.5-sigma procedure; it results in correct statistical evidence that the mixed model distribution does not include that of the observation in slightly more than 28% of the cases compared to the 2.5-sigma procedure's less than 18%. However, if one can tolerate 2 more false positives than the 2.5-sigma procedure, than the lower bound $B_M(0.005;3)$ results in correct statistical evidence that the model distribution does not include that of the observation in slightly over 40% of the cases.

The percentile confidence interval procedure is the most likely to result in correct statistical evidence that the observation distribution is not included in that of the model mixture distribution if the observed time series values tend to always lie above or always lie below the model replications. The 2.5-sigma procedure tends to result in correct statistical evidence that the observation distribution is not included in that of the model mixture distribution when the observed time series tends to lie above the model replications. The 2.5-sigma procedure and percentile confidence interval tend not to result in correct statistical evidence that the model mixture distribution does not include that of the observation when the observed time series tends to lie between the two model distribution modes. The KDE procedure can result in correct statistical evidence that the model distribution does not include that of the observation when the observed time series tends to lie between the two model distribution modes and when the observed time series tends to always lie above or to always lie below the model replications.

Conclusions

The simulation results reported here, and the results of Su et al. (2002) suggest that the currently used 2.5-sigma procedure is unlikely to result in a false positive (incorrect statistical evidence that the mixture model distribution does not include that of the observation when the observation distribution is included in that of the model). However, it is also among the least likely of the procedures considered here to result in correct statistical evidence that the model distribution does not include that of the observation. The efficacy of the KDE procedure depends on the quantile and lower bound chosen. Increased ability to result in correct statistical evidence that the model distribution does not include that of the observation can be associated with an increased chance of incorrect statistical evidence that the model distribution does not include that of the observation when the model mixture distribution does include that of the observation. The simulation results presented here suggest that the KDE procedure with lower bound, $B_M\left(0.001;1\right)$, the maximum number of times a model replication's value KDEs are less than the model replication values' 0.001-quantile or lower bound $B_M\left(0.005;3\right)$ are also unlikely to result in a false positive but are more likely than the 2.5-sigma procedure to result in correct statistical evidence that the mixture model distribution does not include that of the observation.

References

Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*, 863–883. https://www.sas.upenn.edu/~fdiebold/papers/paper16/paper16.pdf

Diebold, F. X., Hahn, J., & Tay, A. S. (1999). Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns on foreign exchange. *The Review of Economics and Statistics*, *81*(4), 661–673.

Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L., & Johnson, N. A. (2008, June). *Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds* (Technical Report No. 537). Department of Statistics, University of Washington.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. URI. http://www.R-project/org/.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall/CRC.

Su, S. Y., McCarty, S. K., Warfield, J. D., Jr., Uthoff, E. J., & Youngblood, S. M. (2022). Evaluation framework for assessing validation methods on modeling and simulation models. *Johns Hopkins APL Technical Digest*, *36*(3), 280–287.

Challenges and Opportunities in Enhancing Department of Defense Ground Vehicle Capabilities through Digital Transformation

Waterloo Tsutsui—is a Senior Research Associate in the School of Aeronautics and Astronautics at Purdue University, Indiana. Tsutsui received his PhD in aeronautics and astronautics from Purdue University in 2017. Before Purdue, Tsutsui practiced engineering in the automotive industry for more than 10 years, with the last position involving the research and development of lithium-ion battery cells for electric vehicles. Tsutsui's research interests are systems engineering, structures and materials, and the scholarship of teaching and learning. Tsutsui is the recipient of the 2023 Engineering Education Excellence Award from the National Society of Professional Engineers (NSPE).

Mikhail Atallah—is the Distinguished Professor of Computer Science at Purdue. A Fellow of ACM and IEEE, his work received Test of Time Awards from CCS and ACSAC and the most prestigious award Purdue bestows in science and engineering (the Bement Award). He was a speaker 10 times in the Distinguished Lecture Series of top CS Departments, and the technology of a startup he co-founded has 5+ billion deployed instances. He was selected as one of the best teachers in the history of Purdue and included in Purdue's Book of Great Teachers, a permanent wall display of its best teachers past and present.

Richard Malak—is an Associate Professor of Mechanical Engineering at Texas A&M University. He holds a BS in electrical engineering from Stony Brook University, an MS in electrical and computer engineering from Carnegie Mellon University, and an MS and PhD in mechanical engineering from Georgia Institute of Technology. He has received two Best Paper and two Paper of Distinction awards by the ASME; at Texas A&M, he has received the Engineering Genesis Award and the Gulf Oil/Thomas A. Deitz Career Development Professorship. Previously, he served as Program Director for the Engineering Design and Systems Engineering program at NSF.

Nathan W. Hartman—is the Dauch Family Professor of Advanced Manufacturing in the School of Engineering Technology at Purdue University, Director of the Digital Enterprise Center, and Director of the Indiana Manufacturing Competitiveness Center (IN-MaC) at West Lafayette. Hartman's research areas focus on the digital transformation of the manufacturing enterprise; tools and methods for model-based definition creation and use in design, production, and sustainment environments; and data interoperability and standards. Funding for his research has come from a variety of sources: NSF, DoD, DoE, the MxD and CyManII federal manufacturing institutes, NIST, and the federally-funded projects for hypersonics and energetic materials at Purdue, as well as industrial partners.

Daniel A. DeLaurentis—is the Vice President for Discovery Park District (DPD) Institutes and the Bruce Reese Professor of Aeronautics and Astronautics at Purdue University. He directs the Center for Integrated Systems in Aerospace (CISA) researching modeling, design optimization and system engineering methods for aerospace systems and systems-of-systems, including urban and regional advanced aerial mobility and hypersonic systems. He is a Senior Research Fellow at the Krach Institute for Tech Diplomacy at Purdue. DeLaurentis served as Chief Scientist of the U.S. DoD's Systems Engineering Research Center (SERC) UARC from 2019 to 2023. He is an elected fellow of the American Institute of Aeronautics and Astronautics (AIAA) and the International Council on Systems Engineering (INCOSE).

Jitesh H. Panchal—is a Professor of Mechanical Engineering at Purdue University. He received his BTech from Indian Institute of Technology (IIT) Guwahati and his MS and PhD from Georgia Institute of Technology. He is a member of the Systems Engineering Research Center (SERC) Council. He is a recipient of NSF CAREER award, Young Engineer Award and three best paper awards from ASME, and was recognized by the Schaefer Outstanding Young Faculty Scholar Award, the Ruth and Joel Spira Award from Purdue University. He is a co-author of two books and a co-editor of one book on systems design.

Abstract

This paper presents valuable insights from in-depth interviews with Department of Defense stakeholders and a rigorous examination of existing guidelines, standards, and pertinent literature. The paper focuses on critical aspects of digital modeling, data utilization, and data-driven decision-making, focusing mainly on the U.S. Army's ground vehicle applications addressing challenges and opportunities. Data-driven decision-making relies significantly on accurate digital twins, which are critical in preparing ground vehicles for their intended environments, especially in challenging environments like preparing vehicles for the Arctic. Thus, creating synergy between real-life applications and digital twins is crucial. However, the U.S. Army faces hurdles in obtaining comprehensive digital data from original equipment manufacturers, especially for older ground vehicle platforms, necessitating reverse engineering to address gaps. Challenges stem from the absence of standardized digital data practices, which triggers the need to establish a cohesive digital modeling framework. To this end, the paper proposes an Intelligent Front-End Framework. The proposed framework optimizes and integrates data management for defense applications and decision-making. To sum up, this paper emphasizes the significance of adopting digital technologies, optimizing and enabling data utilization, and addressing data challenges to enhance the operational readiness and effectiveness of the Department of Defense.

Introduction

The ongoing digital transformation within the Department of Defense (DoD) holds the potential to revolutionize various aspects of its operations, ranging from design and logistics to operations and sustainability. The integration of digital technologies promises substantial improvements in efficiency and effectiveness. Based on a series of interviews with DoD stakeholders, this research dives into the challenges and complexities of this digital transformative journey, mainly focusing on aggregating and incorporating digital models into broader system-level capabilities. While significant progress has occurred in digitization efforts, a critical need exists for a cohesive strategy to ensure that these digital models contribute effectively to mission analysis and optimization through digitalization (i.e., digital transformation).

Our research approach revolved around two core elements: (1) engaging in in-depth discussions with key stakeholders within the DoD and (2) conducting a rigorous examination of existing guidelines, standards, and pertinent literature. For (1), through stakeholder discussions, the authors tap into the wealth of knowledge and expertise possessed by DoD personnel actively involved in the subject matter. Their firsthand perspectives, experiences, and recommendations form a critical foundation for our research. For (2), our comprehensive review process delves into established best practices, industry standards, and the latest advancements in the field. This examination ensures that our research is firmly grounded and up-to-date, allowing us to benchmark our findings against existing frameworks. Combining insights from DoD stakeholders with a review of guidelines and standards, our research approach embodies a holistic, data-driven methodology designed to provide robust and actionable outcomes.

Background

The DoD is at the forefront of bold digital modeling initiatives, aiming to bolster its capabilities across diverse domains. These initiatives encompass a broad spectrum of activities, ranging from the digitization of components via scanning to the creation of intricate 3D models for various vehicle platforms, along with the development of sophisticated simulation models. These coordinated efforts highlight the increasing recognition of the potential benefits of digital technologies, including real-time analysis, predictive modeling, and overall improvements in operational efficiency. However, it is essential to understand that simply creating digital models, while a crucial step, does not guarantee their seamless fit and functionality within the larger framework of the DoD's operations. Despite the abundance of these digital models, significant challenges persist in ensuring their completeness, alignment with mission objectives, and compatibility across various datasets. These challenges are further compounded by the widespread nature of the digitalization process, involving numerous organizations from both the government and the private sector, often spanning international boundaries.

This paper highlights the potential risks to the accuracy of digital models. Unintentional changes, along with deliberate alterations by opposing forces (adversaries) via data hacking, present dangers that could undermine the accuracy and reliability of these models. As a result, creating a robust digital system requires a well-rounded approach. The DoD's digital transformation efforts must go well beyond just the skilled creation of digital data; they must also involve tackling the complex and detailed elements that impact the effective use of digital models across the broader operational setup of the DoD. With these insights in mind, the upcoming sections of this report dive into the specific factors crucial for developing a thorough and unified digitalization strategy.

The distinction between "digitization" and "digitalization" has emerged as a point of critical consideration. To this end, the essential message lies in maintaining data in a format that enables swift content analysis and ensures long-term accessibility through the application of Artificial Intelligence (AI) and Machine Learning (ML) techniques. In this context, "digitization" predominantly involves altering the form of information, while "digitalization" extends further by reshaping both the form and the processes for its creation and utilization. The terms "digitalization" and "digital transformation" share closely related definitions, signifying the profound impact of digital technologies on DoD operations.

Through conducting this research, spanning 2022 to 2023, discussions with DoD representatives have illuminated the changing landscape of the ongoing digital transformation within the DoD. These discussions brought together experts and stakeholders, thereby providing invaluable insights into the challenges deeply embedded in the DoD's digital implementation initiatives. Focused discussions explored digital modeling, data integration, and technological resilience, providing insights into the complex and detailed workings involved in making digital technologies more efficient and effective. The outcomes emphasized the DoD's dedication to fostering a robust digital ecosystem ready to amplify mission success, operational efficiency, and overall readiness in an increasingly complex digital domain.

Current Status of Digital Data and Modeling Infrastructure

Based on the conversation with U.S. Army personnel, the Army acknowledges the potential that digital modeling offers for advancing ground vehicle capabilities. However, this recognition comes with practical challenges that demand careful consideration, particularly in acquiring comprehensive digital data across various vehicle platforms. Within the ground vehicles used by the U.S. Army, the availability of digital data (e.g., geometry data, requirement data, performance data, and analysis data) varies significantly among different vehicle platforms. This variation becomes more apparent when examining vehicles in older platforms (e.g., High Mobility Multi-purpose Wheeled Vehicle [HMMWV], colloquially known as "Humvee"), which are in the process of being replaced by the newer Joint Light Tactical Vehicle (JLTV) Family of Vehicles (FoV; DOT&E, 2011). In the older platforms, the availability of up-to-date digital data is often limited. To address this information gap, the Army must consider reverse engineering. Reverse engineering involves scanning components, processing a point cloud, creating CAD models, and printing the components using a 3D printer via an additive manufacturing process. Therefore, this process involves systematically disassembling and meticulously analyzing physical vehicles to create accurate digital replicas that aim to capture their tangible counterparts' intricate features faithfully.

However, the process of reverse engineering, though methodical, has challenges and potential limitations. Developing precise digital representations requires a deep knowledge of each vehicle's design and construction details. Despite these dedicated efforts, a drawback emerges: digital renditions may unintentionally overlook specific subtle characteristics and variations inherent to their tangible counterparts. Also, the utilized/original components for generating the digital representation may include notable manufacturing discrepancies (errors) and could fall outside acceptable tolerances. Consequently, the resulting digital models may not faithfully portray the initial design intent. Furthermore, reverse engineering typically captures shape without capturing behavioral, performance, or contextual parameters that interplay with the geometry. These parameters ensure that the digital models accurately simulate real-world conditions and behaviors.

Moreover, a critical consideration arises: Should we also capture the test/qualification processes that the engineering teams underwent to qualify these parts? This concern arises from the possibility that a 3D-printed version will not meet specifications or performance requirements. In order to address this concern, our research aims to determine what needs to be captured in this regard, recognizing that comprehensive data on the testing and qualification processes is crucial for validating the accuracy and functionality of the digital models and their real-world applications.

As a result of the complicated engineering tasks involving reverse engineering, a significant challenge arises: the lack of a streamlined digital data acquisition process and comprehensive data repositories. This absence introduces a range of complexities when constructing digital models for the diverse ranges of ground vehicles. The absence of standardized digital data further amplifies the challenges in establishing a cohesive digital modeling infrastructure (i.e., a well-organized and interconnected system for creating digital models), which is essential for fully leveraging the potential of digital modeling within the operational domain of the U.S. Army.

Figure 1 illustrates the development of a jet engine in a streamlined circular and iterative digitalization process. Although primarily associated with aircraft, it shares essential principles with ground vehicle development, including those designed and manufactured for the U.S. Army. While the specific applications differ, the fundamental approach to systems engineering, digitalization, and iterative processes remains similar. Both disciplines involve deconstructing complex systems into interconnected components, iterative design refinements, adaptation to evolving requirements, rigorous testing, risk mitigation, compliance with safety standards, and environmental considerations. These shared principles highlight the fundamental principles of systems engineering with a circular and iterative digitalization process, making it a valuable digitalization framework for innovation and efficiency across various engineering domains, including aerospace and ground vehicle development.

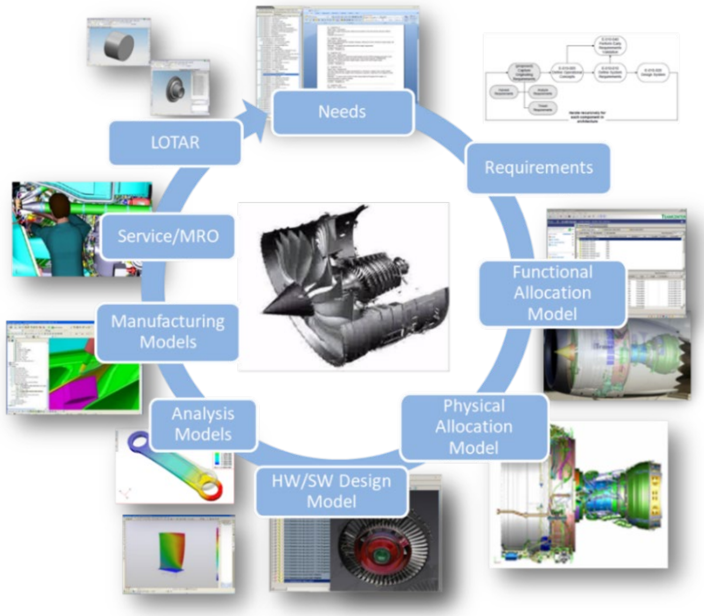


Figure 1. Iterative Circular Digitalization Process in Systems Engineering

Figure 2 examines the model-based digital product data (Hartman & Zahner, 2017) connecting the other information domains in an enterprise setting. The digital product data (Figure 27, lower bottom) is a critical bridge connecting multiple data domains within an organization, fostering streamlined operations and informed decision-making. It supports analytics and insights (Figure 2, top left) by providing the foundational data for uncovering patterns, anomalies, and actionable insights that inform product optimization and strategic decisions. Moreover, digital product data influences user interfaces and experiences, ensuring that interfaces are responsive and intuitive, enhancing overall user satisfaction. Data collection and integration (Figure 2, middle) act as a central repository, consolidating information from various product lifecycle stages and enabling seamless team collaboration.

Research shows that digitalization in the ground vehicle value chain offers rich research opportunities in all connected fields throughout the enterprise, including supply chain management, digital twins, virtual prototyping, and AI for optimization and predictive analytics (Panchal & Wang, 2023). Lastly, the digital product data extends its reach across the entire enterprise (Figure 2, right), enhancing organizational capability by ensuring all departments have access to the latest product information. This interconnectedness empowers organizations to operate efficiently, innovate effectively, and maintain competitiveness.

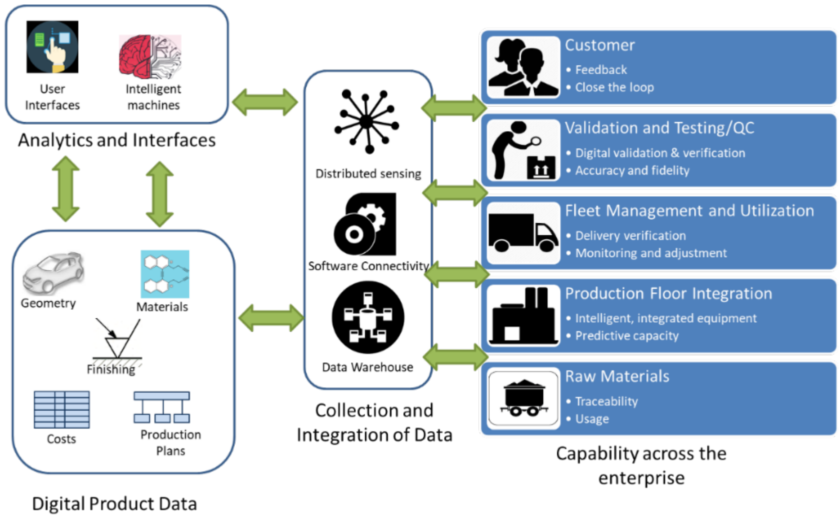


Figure 17. Digital Product Data Integration and Enterprise Connectivity Adapted from Kinnet (2015)

Figure 3 examines the data exchange between physical and virtual vehicle models, a critical element of digitalization in product design and development. This data exchange streamlines communication and synchronization between the physical prototype (Figure 3, left) and its digital counterpart (Figure 3, right). It increases accuracy through precise measurements, iterative refinements, and comprehensive simulations, fostering early issue detection and cost savings. This data exchange significantly shapes decision-making throughout the design and development process, providing real-time feedback, facilitating iterative refinements, and enabling comprehensive evaluation of design alternatives. Consequently, decision-makers benefit from a wealth of insights, leading to more informed and cost-effective choices in the quest for optimal vehicle design and development.

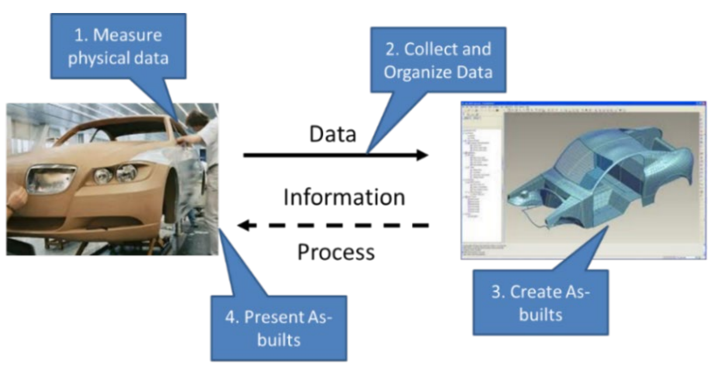


Figure3. Data Exchange Between Physical Model and Virtual Model

Figure 4 demonstrates the versatile application of digital models across a product's lifecycle, spanning from the "As Designed" to "As Manufactured" and "As Used" stages. This concept parallels their role in bridging the gap between these states, exemplified by jet engine blades (similar to the aircraft component example in Figure 1) and maintaining a consistent engineering philosophy with ground vehicles utilized by the U.S. Army. A comprehensive digital representation initially captures the blade's intended design, serving as the foundation for anticipated performance with minimal variability. During manufacturing, rigorous inspections ensure precise conformance with design specifications, fostering robust quality control.

However, inherent manufacturing variability introduces distinctions between "As Designed" and "As Manufactured." Over the product's operational life, real-world usage data seamlessly integrates with the digital model, empowering engineers to monitor performance, promptly identify deviations, and enable predictive maintenance vigilantly. This variability assumes particular significance under "As Used" conditions, where mechanical and thermal stresses subject the product to rigorous trials. These advanced digital models are indispensable tools, elevating product performance and efficiency, fortifying reliability, and upholding stringent safety standards through meticulous data-driven optimization practices.

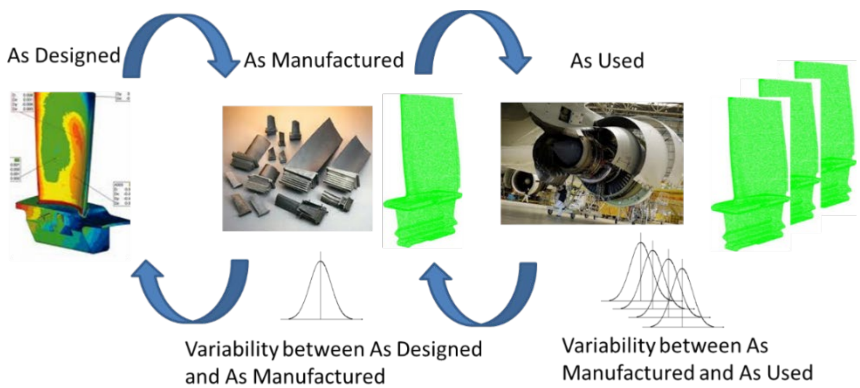


Figure 4. Comparison of Product: As Designed vs. As Manufactured vs. As Used

The U.S. Army's recognition of the promise offered by digital modeling in the context of ground vehicles contrasts with the practical realities stemming from varying degrees of digital data availability, complexities of reverse engineering, and the lack of standardized data repositories. While capturing the digital representation of legacy weapons platforms will involve much manual work and effort, it is not without benefits. It will simply be a choice the Army will need to make in light of concerns around intellectual property (IP), budget, and readiness. Speaking of IP, contracts with Original Equipment Manufacturers (OEMs) should include some level of access to digital products and process data to support sustainment over the lifecycle. In the ongoing balance between goals and limitations, the U.S. Army is grappling with various difficulties while striving to build a robust digital modeling infrastructure. However, as the U.S. Army implements robust infrastructure for digital modeling, it will markedly elevate its operational capabilities and broaden its scope of accomplishments.

Digital Modeling for Ground Vehicles

Within the evolving landscape of ground vehicle capabilities, a central focus emerges on the transformative potential of digital modeling. This discussion notably underscores the crucial role that digital modeling plays in understanding the complex nature of ground vehicle performance, especially in demanding environmental conditions. Furthermore, the synergy between non-destructive testing (NDT) and digital modeling is essential. NDT, through its non-invasive data acquisition, complements digital modeling capabilities. While NDT provides real-world insights, digital modeling excels in analyzing, simulating, and predicting performance under diverse conditions. This harmonious collaboration empowers engineers and decision-makers to make well-informed choices in design, maintenance, and optimization, thereby enhancing the reliability and safety of ground vehicles across a spectrum of operational scenarios.

Central to this discussion is the strategic goal of utilizing digital models for simulations that forecast outcomes like those seen in previous research, where researchers assessed model-based, decision-support frameworks (Malak et al., 2009; Tsutsui, Guariniello, et al., 2023; Tsutsui, Shi, et al., 2023). The overarching goal is to equip decision-makers with a proactive understanding of ground vehicle behavior under various environmental conditions, enabling them to anticipate and strategize for diverse vehicle use conditions. This forward-looking approach empowers the U.S. Army to make well-informed decisions that optimize ground vehicle configurations and operational strategies, thereby enhancing their efficiency and effectiveness in traversing the demanding terrains in harsh environments.

Venturing beyond traditional boundaries, this research discussion introduced thought-provoking inquiries into the untapped potential of digital models. Exploring scenarios that go beyond the norm opens pathways to adaptability and customization. By capitalizing on the flexibility offered by digital modeling, decision-makers unlock opportunities to tailor ground vehicles to meet the specific operational demands of challenging environments. This adaptable methodology ensures that ground vehicles remain finely tuned to navigate the distinct challenges presented by these harsh landscapes.

This ongoing conversation underscores the profound significance of digital modeling in unveiling ground vehicle capabilities. Through the strategic deployment of digital models for predictive insights and the exploration of innovative scenarios, the U.S. Army effectively positions itself at the forefront of harnessing technological advancements to address the complexities of operating vehicles in the most challenging environments, like those in the Arctic and beyond.

Ground Vehicle Requirements for the Arctic Environment

In May 2022, the Army announced its plan to establish a new Alaska-based division, the 11th Airborne Division (Congressional Research Service, 2023). This initiative involves activating new units and reconfiguring two existing Alaskan Infantry Brigade Combat Teams (IBCTs). As this new division takes shape and additional forces take their positions in Alaska, the dynamics of ground vehicle requirements, specifically concerning the JLTVs, could potentially undergo significant changes.

Engaging in conversations with Army personnel highlights a significant focus on the crucial task of adapting ground vehicles to excel in the challenging conditions of Arctic environments, where the vehicle operating temperature can plummet as low as -50 degrees Fahrenheit (Eversden, 2022; Zielinski & Maguire, 2022), while another source (Rozell, 2011) even suggests that temperature may drop to as low as -80 degrees Fahrenheit. Characterized by their extremely cold temperatures, these landscapes demand vehicles that meet functional requirements and demonstrate resilience in harsh climatic conditions. Therefore, examining ground vehicle requirements for the Arctic becomes a vital consideration, emphasizing the essential role of digital modeling in guiding design and decision-making processes.

At the core of these discussions lies Army Regulation (AR 70-38, 2020), as shown in Figure 5, Army Techniques Publication (ATP 3-90.97, 2016), DoD Comprehensive Selected Acquisition Report (SAR) on JLTV (DAMIR, 2019), and Special Report (Walsh et al., 1989). The reference to the AR document (AR 70-38, 2020) emphasizes the establishment of operational benchmarks based on diverse climate classifications. At the same time, the detailed conversations further emphasized the significant role of temperature-related considerations in guiding the formulation of design attributes and performance benchmarks for ground vehicles well-suited for Arctic terrains.

High Temperature
- Differential expansion of dissimilar materials
- Increased rate of chemical reactions
- Lower lubricant viscosity
- Electronic circuit instability
- Outgassing

Rain
- Moisture in parts
- Corrosion
- Attenuation of electromagnetic radiation
- Reduced off-road mobility

Icing and Snow
- Reduced visibility
- Attenuates electromagnetic signals

High Humidity
- Triggers Condensation
- Initiates fogging
- Oxidation/corrosion
- Increased chemical reaction
- Changes in material properties
- Degrades optical and infrared properties
- Modifies lubricants' behavior
- Changes elasticity/plasticity

Low Temperature
- Changes physical properties of materials
- Hardness and brittleness
- Differential contraction
- Increased viscosity
- Stiffening of shock mounts
- Condensation and freezing of water

High Elevation (low pressure)
- Liquids leak
- Lubrication capacity decreases
- Electrical breakdown
- Heat transfer is less efficient
- Liquids vaporize
- Less efficient combustion

Salt/Saltfog
- Acceleration of metallic corrosion
- Loss of mechanical and structural strength
- Surface deterioration
- Alteration of electric properties

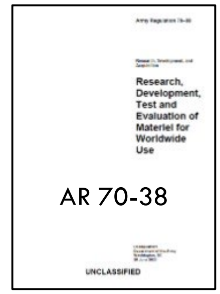


Figure 5. Effects of Different Environmental Conditions (AR 70-38, 2020)

These comprehensive guides establish essential criteria for effective ground vehicle operations and equipment affected by the distinctive challenges posed by the Arctic's harsh environments (Table 1). These documents' guidelines lay the foundation for formulating design principles and operational strategies, ensuring ground vehicles' adaptability and robustness to perform effectively in the harshest cold conditions.

Furthermore, the discussions with U.S. Army personnel revealed a key emphasis on the complex relationship between temperature, vehicle performance, materials, and components. This complex mix of factors shapes the detailed considerations behind crafting ground vehicles suited for Arctic missions. The significant impact of temperature (i.e., thermal loading) on various vehicle aspects underscores its importance as a critical factor influencing the fundamental characteristics of vehicles designed for Arctic use. This understanding calls for a comprehensive design approach that covers vehicle dynamics and thoughtfully incorporates the Arctic environment's distinct thermal dynamics.

Table 3. Equipment Affected by Arctic Environment (Walsh et al., 1989)

Assembly	Mechanical	Electrical	Fluid	Structural
Power source	-	Battery, cables, Fuse, Connector	Acid	Mount
Starter system	Switch, belt, bearings	Glow plugs, cables, fuses, relays, alternator	Lubricants	Mount
Fuel system	O-rings, filter, seals, fuel lines, bearings, valves	Fuel pressure sensor, solenoid valves, motor	Fuel	Fuel tank
Engine	Rings, bearings, gaskets, O-rings, springs, chain drive, seals, valve guides	Glow plugs	Sealants	Engine mounts
Lubrication system	Seals, o-rings, bearings, piping, filters, connectors	Pressure sensors, level sensors, temperature sensors	Lubricants	-
Exhaust	Fasteners	Exhaust gas pressure sensor	-	-
Controls	Switches, governor	Switches, meters, connectors, voltage regulator	-	-
Gearbox/clutch	Bearings, seals, spring, gasket, coupling	-	-	-
Generator	Bearings, seals, brushes, springs, coupling	Wiring, coil, stator, connectors, field detector	-	Mounts
Frame mount	Fasteners, bushings	Grounding strap	-	Frame anchors, mounts
Cooling system	Seals, bearings, hoses, fan, radiator	Wiring, temperature sensors, liquid level sensors	Coolant	Mounts

Given these recent developments, the Army's establishment of the 11th Airborne Division in Alaska will most likely introduce an additional aspect to the domain of ground vehicle requirements. With new division units, reconfigured IBCTs, and increased forces in the Arctic region, the evolving demands for JLTVs may experience significant adjustments. As the Army's strategic posture adapts to this changing context, integrating ground vehicle requirements with the evolving divisional structure remains critical, shaping the path toward optimized operational effectiveness and strategic agility.

Data-Driven Decision-Making for Ground Vehicle Selection and Preparation

The pursuit of data-driven decision-making takes center stage when selecting and readying ground vehicles, especially in unique environments like the Arctic. The objective is to leverage the insights from extensive data to inform and guide the streamlined process of choosing and preparing vehicles that meet operational requirements and exhibit resilience in extreme conditions.

A comprehensive array of data attributes facilitates informed decisions for U.S. Army personnel. Considerations span a broad spectrum, encompassing critical factors such as mobility, tire selection, powertrain specifications, oil/fuel selection, structural attributes, shock absorption capabilities, reliability metrics, gradeability performance, and other relevant attributes (AR 70-38, 2020). This diverse range of data collectively serves as the bedrock upon which experts meticulously craft robust and informed decisions, ensuring that the selected ground vehicles are fit for purpose and optimized for the specific challenges of Arctic environments.

A critical focus that emerged from the discussions with the U.S. Army personnel was the detailed understanding that different vehicle types and platforms require varying levels of data detail. Recognizing this diversity underscores the significance of tailoring data acquisition efforts to the specific requirements and complexities of each distinct vehicle type and platform. Furthermore, the spotlight turns to the essential role of accessing OEM data, an indispensable resource that serves as a cornerstone for effective decision-making. Accessing OEM data becomes a valuable asset, offering a reliable and authoritative foundation for data-driven choices.

Introducing a potential disruption, the change of the supplier for JLTVs from Oshkosh to AM General (Tadjdeh, 2022; Tricomo, 2023) raises concerns regarding access to OEM data. This shift could impact the availability of crucial data required for data-driven decision-making in ground vehicle selection and preparation. The transition between suppliers may introduce complexities in obtaining OEM data, potentially hindering the seamless and informed decision-making process that relies on comprehensive and accurate OEM data. Since the JLTV supplier is changing from Oshkosh to AM General (Magnuson, 2023), addressing the potential implications on data access and ownership becomes essential in maintaining the decision-making framework's robustness.

Exploring data ownership and access further, the conversation with the U.S. Army personnel unveiled a detailed examination of contrasting approaches: owning data versus accessing data. The discussion dived into the complexities of data rights, ownership, and their implications for informed decision-making. Challenges often stem from the complexities of data access and ownership agreements. The dialogue emphasizes the importance of balancing owning and accessing data, recognizing the vital roles in cultivating a comprehensive and knowledgeable approach to selecting and preparing ground vehicles.

In short, the discussion about requirement-driven decision-making for ground vehicle selection and preparation reflected a solid commitment to using the requirements specified in the reference document (AR 70-38, 2020) as a guide. These discussions encompassed a wide range of aspects, carefully chosen to ensure optimal performance and adaptability of ground vehicles in the challenging Arctic environment. The discussion also emphasized the importance of accessing OEM data while addressing data ownership and access challenges, resulting in a comprehensive framework that empowers decision-makers to confidently tackle the challenges of vehicle selection and preparation with precision.

Analyzing the Potential of Integrated Digital Transformation

Conversations with U.S. Army personnel offered valuable insights into ground vehicle modeling and preparation within the DoD scope. These discussions, centered on the challenges and possibilities associated with ground vehicles in extreme environments, unveil compelling opportunities for leveraging mission-aware integrated digital transformation to maximize the operational advantage of the DoD.

Central to these conversations is the significant role of digital modeling, which emerges as a powerful tool for enhancing the decision-making process. By creating accurate digital replicas (i.e., digital twins) of ground vehicles, the DoD gains the capacity to simulate and evaluate vehicle performance under various conditions. In scenarios like the Arctic, known for its extreme cold, this simulation-driven approach proves invaluable for well-informed decision-making about vehicle selection, operation, and strategic adaptations.

Moreover, it is worth noting that these digital twins' accuracy hinges on maintaining a comprehensive per-vehicle part number database throughout the sustainment process. Within this context, the comprehensive per-vehicle part number database should include information about the specific software versions used in each vehicle. In the context of digital twins and ground vehicles, software versions can be critical because they may impact how the vehicle operates, its capabilities, and its compatibility with other systems. Including information about the software versions in the database ensures that the digital twin accurately represents the vehicle's configuration and capabilities, which is essential for effective simulation and decision-making. Maintaining a real-time data stream of the vehicle's operating environment is crucial. While much of the discussion centers around comprehending the thermal environment, the scope extends beyond that. Marine/saltwater exposure and humidity can also accelerate material degradation. Therefore, the effectiveness of the per-vehicle twin relies on its ability to comprehensively understand the actual operating environment, ensuring a more accurate basis for decision-making.

Further emphasizing this potential is the ongoing discussion surrounding ground vehicle requirements in extreme settings. Here, the necessity of operational adaptability comes to the forefront. The digital twins equip the DoD to analyze and tailor vehicles for non-standard conditions rapidly. By simulating the effects of extreme temperatures on vehicle components and materials, the DoD can proactively address potential performance constraints and seamlessly optimize vehicle designs to align with the unique demands of specific environments. However, the path to realizing these benefits is not without its challenges. The dialogue highlights complexities related to data accessibility and ownership, particularly concerning acquiring digital data from OEMs. In order to navigate this terrain, mission-aware digital integration emerges as a strategic avenue for establishing collaborative partnerships and securing timely access to crucial data. The DoD can construct a comprehensive overview of ground vehicle capabilities and limitations by gathering data from various sources.

Beyond its role in decision-making, integrating digital models and data across varied ground vehicle platforms holds promise for streamlining logistics and maintenance processes. Leveraging these digital models allows the DoD to anticipate maintenance needs, optimize supply chains, and devise strategies for vehicle repairs or component replacements. This proactive approach minimizes operational downtime while maximizing operational readiness, regardless of the vehicles' operation locations. Furthermore, the discussion highlights how digital models can undergo customization for different mission profiles and scenarios. With the ongoing discussion of mission-aware integrated digital transformation, the DoD can create specific simulations that accurately mirror various operational scenarios' unique challenges and specific needs. This tailored simulation capability enhances the accuracy and relevance of decision-making processes.

Collecting these insights points to the digital implementation effort to boost the DoD's operational advantage. By utilizing digital modeling, simulations, and data-driven decision-making, the DoD can skillfully adapt ground vehicles for various environments, improve mission readiness, and maximize resources. Tackling data access and ownership complexities is crucial, as it holds the key to unlocking the full range of possibilities in integrated digital transformation. As the DoD tackles these challenges, the benefits become apparent through enhanced operational excellence.

Test, Evaluation, and Data Throughout the Acquisition Process

The DoD employs a structured five-step process for Testing and Evaluation (T&E), as depicted in Figure 6 (Johnson et al., 2005). In Step 1, stakeholders identify T&E information, focusing on their needs and evaluating systems across various stages, from concepts to production systems. In Step 2, stakeholders scrutinize evaluation objectives during the pre-test analysis phase, determine data types, and design test scenarios using validated models and simulations. In Step 3, tests are planned and executed during test activity and data management, focusing on gathering accurate and complete data while assessing historical data availability. In Step 4, post-test synthesis and evaluation involve comparing measured outcomes with expected ones, identifying deviations, and utilizing Modeling and Simulation (M&S) techniques to analyze data meaningfully. Finally, in Step 5, decision-makers use the synthesized T&E information and other factors to determine the course of action, potentially leading to further iterations of the DoD T&E process to refine system capabilities.

It is worth noting that the structured T&E process outlined by the DoD (Johnson et al., 2005) will need to continually transform itself for better verification and validation practices when considering the fact that we move toward a more agile acquisition process (DeLaurentis et al., 2022). In order to deliver products and capabilities faster to meet rapidly evolving defense needs through a more agile acquisition process, we will need to integrate digital twins and move the five steps faster and more efficiently.

Regarding testing and evaluation standards for military equipment's environmental resilience, NATO AECTP 300 (2006), *Climatic Environmental Tests*, establish a comprehensive frameworkClick or tap here to enter text.. These standards, devised by NATO, aim to assess the endurance of military gear and equipment across a spectrum of weather conditions. The evaluations ensure the equipment can withstand environmental stresses throughout its operational life. The NATO testing regimen encompasses various specific tests, as shown in Table 2.

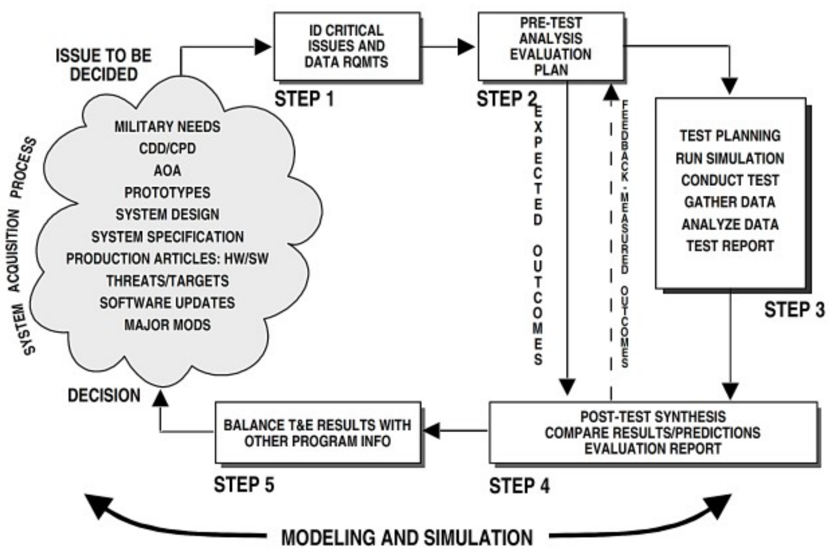


Figure 6. Testing and Evaluation Process
(Johnson et al., 2005)

Table 4. Climate Environmental Tests
(NATO AECTP 300, 2006)

Method	Title
301	General Requirements
302	High Temperature
303	Low Temperature
304	Thermal Shock
305	Solar Radiation
306	Humid Heat
307	Immersion
308	Mould Growth
309	Salt Fog
310	Rain and Water Tightness

Method	Title
311	Icing
312	Low Pressure
313	Sand and Dust
314	Contamination by Fluids
315	Freeze and Thaw
316	Explosive Atmosphere
317	Temperature, Humidity and Altitude
318	Vibration, Temperature, Humidity, and Altitude
319	Acidic Atmosphere lifespan

Exploring further into the discussion with the U.S. Army personnel revealed a practical approach to utilizing specific documents in the informed decision-making process. The authors learned that the following key steps are crucial guides in this process: the Initial Capability Document (ICD), Capability Design Document (CDD), and Capability Production Document (CPD). Figure 7 depicts the steps involving the ICD, CDD, and CPD. The ICD serves to identify essential mission capabilities, assess gaps and their priorities, recognize operational risks, and highlight the imperative to address these gaps. Meanwhile, the CDD plays a vital role in confirming Key Performance Parameters (KPP); describing their thresholds and goals; evaluating costs, schedule, and technology risks; and assessing system affordability vis-à-vis the delivered operational capability. KPP types for JLTV include mobility, transportability, survival, payload, sustainment, net-ready, and system training (DAMIR, 2019). Lastly, the CPD ensures that the delivered system aligns with the initially defined needs outlined in the ICD. This comprehensive approach underscores the importance of meticulous planning and assessment in military operations.

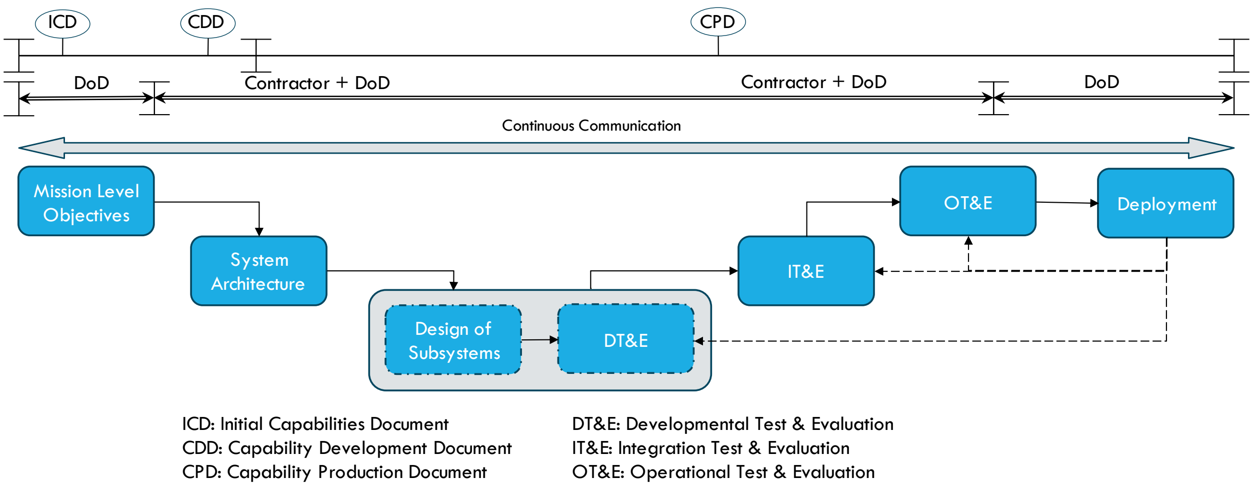


Figure 7. Data Generated Throughout the Acquisition Process

Enhancing Data Utilization and Institutional Memory With an Intelligent Front-End (IFE)

Building upon insights gained from conversations with DoD stakeholders and emphasizing the crucial role of efficient data utilization, the authors aim to transition the conversation from simply understanding the current landscape to presenting a comprehensive framework that elevates both data utilization and institutional retention. Introducing the Intelligent Front-End (IFE) framework, this section dives into optimizing data management, integration, and utilization. By developing a workforce for the digital enterprise (Hartman et al., 2019), empowering decision-makers, and reinforcing the DoD's institutional memory, the IFE framework aligns seamlessly with our ongoing mission of achieving effective digital transformation.

Efficient use of data cannot be just "nice to have" in domains like military operations, equipment maintenance, and decision-making. It is a must. When people interact with data, they unintentionally send signals about what matters. However, making the most of these signals requires an innovative approach to turn them into concrete improvements. This is where the IFE comes into play. As a bridge between existing systems and modern data needs, the IFE decodes these signals and transforms them into practical enhancements, blending human insights with technological advancements.

The innovative approach behind the IFE dives into the domain of user signals. IFE elevates data delivery, making it precise and exceptionally user-friendly. Thus, users can use this system as a reliable data partner, thereby helping users meet the demand for accurate, adaptable, and user-centric data utilization. With the IFE constantly learning, improving, and fine-tuning responses, it is like having a knowledgeable data ally. This carries significant implications in various contexts, from informed decision-making to streamlined operations. All these pieces come together, and the IFE becomes a transformative force across the spectrum. In the following section, the description will dive into the interactions among users, the existing/legacy systems, and the IFE within the context of three distinct phases (Phases 1–3).

In Phase 1 (Figure 8), "Learning," the IFE operates discreetly in the background as an intermediary between users and the currently used/legacy systems. The system transmits user queries (Q) to the currently used systems and promptly relays the resulting responses (R) back to the users. During this phase, the IFE takes on the role of an attentive observer, closely studying the interactions between users and the currently used/legacy systems. This process of observation and learning helps the IFE gain insights into user preferences, patterns, and the effectiveness of responses. Additionally, users have the option to provide quick feedback (i.e., User rating of R) by giving a simple one-click rating to indicate how useful they found the provided response.

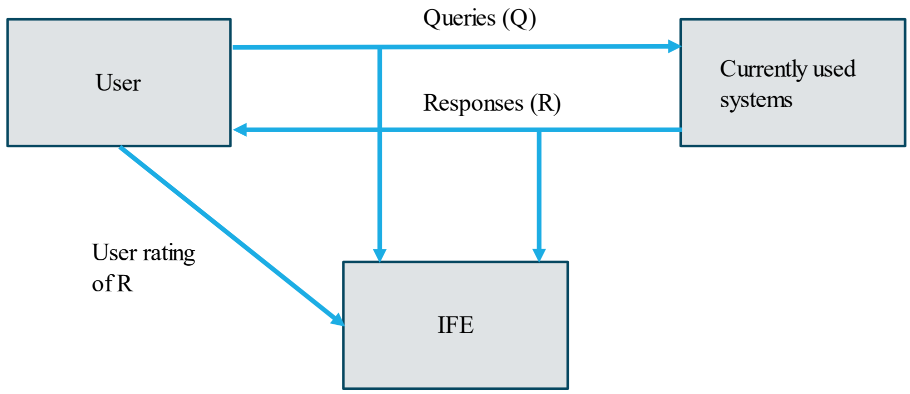


Figure 8. Phase 1: Learning (IFE is Passive)

In Phase 2 (Figure 9), "Dual Deployment," users are given a dual-choice option. They can either directly engage with the familiar legacy systems as they have done in the past (Figure 8), or they can opt for an alternative route by using the IFE. This is similar to taking the traditional road or exploring a new, more optimized path. During this phase, the IFE becomes more actively involved. The IFE steps in to optimize the communication process by adjusting both the user queries (Q') and the responses received (R'). This proactive approach ensures that the delivered information is finely tuned to meet the specific needs and expectations of the users. This dual deployment lasts until the new IFE improves over the old IFE (i.e., User rates R vs. R").

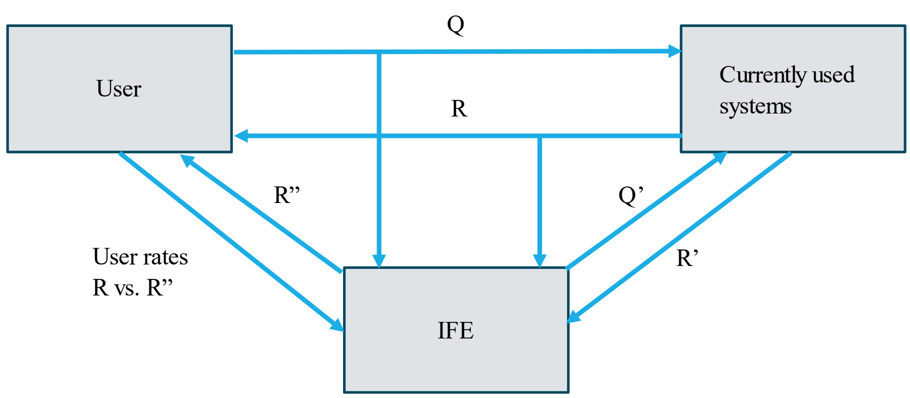


Figure 9. Phase 2: Dual Deployment (IFE is Active, but It Is Not Fully in Charge)

In Phase 3 (Figure 10), "Fully Deployed IFE," the IFE takes a central position, directly aligning itself between the user and the currently used/legacy system, thereby eliminating the need for dual deployment. All information and interactions flow exclusively through the new IFE, which serves as the primary conduit for user queries, responses, and data communication. This streamlined configuration ensures a unified and optimized user experience, where the IFE seamlessly facilitates data exchange while intelligently enhancing the interaction process.

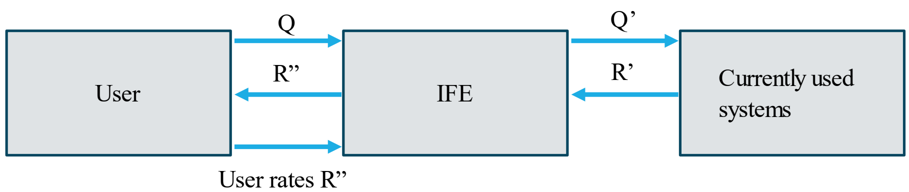


Figure 10. Fully Deployed IFE

Integrating the innovative approach using the IFE enhances data utilization and contributes to establishing and preserving institutional memory through digital technology. The IFE’s perceptive understanding of data utilization patterns and user interactions plays a central role in accumulating valuable organizational insights. As users engage with data, the IFE captures these interactions, gradually constructing a digital repository of institutional knowledge. This synergistic relationship between refined data utilization and institutional memory enhances decision-making processes and facilitates the seamless transfer of organizational knowledge and expertise. This proposed approach ensures operational continuity and adaptive responses to dynamic challenges. Embracing the innovative approach of using the IFE highlights the potential to elevate data utilization, improve operational efficiency, and strengthen institutional memory’s core basis/foundation, collectively shaping a transformative landscape for data-driven activities.

Conclusions

The conversations with DoD stakeholders have highlighted critical aspects of digital modeling, ground vehicle preparation, and data-driven decision-making within the U.S. Army. These discussions have revealed the potential benefits and challenges associated with leveraging digital technology to enhance the DoD’s operational advantage. One key takeaway is recognizing the transformative potential of digital modeling in understanding and optimizing ground vehicle performance, particularly in extreme environments like the Arctic. The creation of accurate digital twins provides a powerful tool for simulating and evaluating vehicle behavior under diverse conditions, facilitating well-informed decision-making in vehicle selection, operation, and adaptation.

However, realizing these benefits entails navigating complexities. We must address challenges related to data accessibility, ownership, and the transition between vehicle suppliers. In order to navigate these challenges, a mission-aware integrated digital transformation approach is crucial. This approach involves collaborative partnerships, data integration, and tailored simulations to enhance decision-making accuracy and operational readiness. Furthermore, the discussions have highlighted the importance of maintaining comprehensive per-vehicle part number databases and real-time data streams to ensure the accuracy of digital twins. In order to support effective decision-making, these digital representations must encompass the vehicle’s physical attributes, software versions, and the actual operating environment.

In addition to decision-making, integrating digital models and data holds promise for streamlining logistics, maintenance processes, and supply chain optimization. This proactive approach minimizes operational downtime and maximizes readiness. The DoD’s journey toward digital transformation presents a wealth of opportunities to enhance operational excellence. By embracing digital modeling, simulations, and data-driven decision-making, the DoD can adapt ground vehicles for diverse environments, improve readiness, and maximize resources.

In our future plan, we aim to create a versatile decision/reasoning tool framework tailored to various cases, providing high-level guidance to sponsors based on crucial decision-making factors. This framework will allow us to prioritize modeling efforts to address specific decision needs. Additionally, when collaborating with a DoD unit, we intend to move toward obtaining clearance-required information and gaining insights into their decision-making processes. To this end, the research team would like to investigate the possibility of formalizing a DD 254 to specify the classification requirements for contracts with relevant entities to ensure access to essential files (e.g., ICD, CDD, and CPD) for future research. Finally, through these future initiatives, our goals are to refine our approach continuously, leveraging digital modeling and data-driven decision-making to meet evolving sponsor requirements, thereby assisting the DoD in enhancing its operational advantage.

Acknowledgments

The authors acknowledge financial support from the U.S. Department of Defense through SERC/AIRC on research task WRT 1057.18d, contract no. HQ0034-19-D-0003 and report no. AIRC-2023-TR-013. The authors are immensely grateful to Mr. Steve McKee for his vital support in advancing our research. The authors recognize the invaluable contributions of Ms. Nickee Abbott, Mr. Harry Bailey, Mr. Bill Baker, Mr. James Colson, Ms. Jazmine Garard, Mr. Sanjay Kankanalapalli, Dr. Sebastian Karwaczynski, Mr. Eric Linderman, Dr. Dennis McBride, Mr. Joseph Sparks, Mr. Paul Strzalkowski, Mr. Mark Temnycky, Mr. Van Weaver, and Mr. Richard Wimberly, who participated in interviews and meetings to discuss DoD practices in digital transformation. Lastly, the authors thank the leaders of SERC/AIRC—Dr. Dinesh Verma, Dr. Philip Anton, Dr. Douglas Buettner, Ms. Kara Pepe, and Ms. Tara Kelly—for their exceptional management of the funded project, facilitation of the DoD-university partnerships, and assistance in transitioning academic research into practical solutions that benefit the DoD and our dedicated warfighters.

References

AR 70-38. (2020). *Army Regulation 70–38, research, development, test and evaluation of materiel for worldwide use*. https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN30017-AR_70-38-000-WEB-1.pdf

ATP 3-90.97. (2016). In *Army Techniques Publication (ATP), Mountain Warfare and Cold Weather Operations*.

Congressional Research Service. (2023). *Congressional Research Service (CRS) Reports, joint light tactical vehicle (JLTV)*. https://crsreports.congress.gov/product/pdf/IF/IF11729/6

DAMIR. (2019). *Defense Acquisition Management Information Retrieval (DAMIR), joint light tactical vehicle*. https://www.esd.whs.mil/Portals/54/Documents/FOID/Reading Room/Selected_Acquisition_Reports/FY_2019_SARS/20-F-0568_DOC_47_JLTV_SAR_Dec_2019_Full.pdf

DeLaurentis, D. A., Moolchandani, K., & Guariniello, C. (2022). *System of systems modeling and analysis*. CRC Press.

DOT&E. (2011). *Director Operational Test and Evaluation, FY11 Army programs, joint lightweight tactical vehicle (JLTV)*. https://www.dote.osd.mil/Portals/97/pub/reports/FY2011/army/2011jltv.pdf?ver=2019-08-22-112309-190

Eversden, A. (2022). *BAE, Oshkosh cold weather vehicle prototypes survive Army's Alaskan tests*. https://breakingdefense.com/2022/01/bae-oshkosh-cold-weather-vehicle-prototypes-survive-armys-alaskan-tests/

Hartman, N. W., Herron, J., Astheimer, R., Fuerst, T., & Hess, D. (2019). A need for digital enterprise workforce development. *Model-Based Enterprise Summit (MBE 2019)*, 64.

Hartman, N. W., & Zahner, J. (2017). Extending and evaluating the model-based product definition. *NIST GCR 18-015*.

Johnson, C., Cavoli, C., & Claxton, J. D. (2005). *Test and evaluation management guide*. Defense Acquisition University.

Kinnet, J. (2015). *Creating a digital supply chain: Monsanto's journey*.

Magnuson, S. (2023). AM General wins JLTV re-compete contract over Oshkosh. *National Defense*.

Malak, R. J., Jr., Aughenbaugh, J. M., & Paredis, C. J. J. (2009). Multi-attribute utility analysis in set-based conceptual design. *Computer-Aided Design*, *41*(3), 214–227.

NATO AECTP 300. (2006, January). *Climatic environmental tests*. Brussels, Belgium.

Panchal, J. H., & Wang, Z. (2023). Design of next-generation automotive systems: Challenges and research opportunities. *Journal of Computing and Information Science in Engineering*, *23*(6).

Rozell, N. (2011). *Alaska's all-time cold record turns 40*. Geophysical Institute, the University of Alaska Fairbanks. https://www.gi.alaska.edu/alaska-science-forum/alaskas-all-time-cold-record-turns-40

Tadjdeh, Y. (2022). Recompete for JLTV offers coveted prize for vehicle makers. *National Defense*. https://www.nationaldefensemagazine.org/articles/2022/2/14/recompete-for-jltv-offers-coveted-prize-for-vehicle-makers

Tricomo, S. (2023). Army announces joint light tactical vehicle follow-on production award. In *U.S. Army*.

Tsutsui, W., Guariniello, C., Mall, K., Patterson, F., Balestrini-Robinson, S., Panchal, J., & DeLaurentis, D. (2023). *Model-based approach in defense portfolio management: Data preparation, analysis, and visualization of decision spaces*.

Tsutsui, W., Shi, Q. A., Walter, I., Wei, A., Williams, C., DeLaurentis, D., & Panchal, J. (2023). Decision making for additive manufacturing in sustainable defense acquisition. *Naval Engineers Journal*, *135*(4), 47–58.

Walsh, M. R., Morse, J. S., & others. (1989). *Preliminary design guide for arctic equipment*. SR 89-13. US Army Corps of Engineers - Cold Regions Research and Engineering Laboratory.

Zielinski, S., & Maguire, P. (2022). Hot and cold. *Acquisition, Army ALT Magazine, Science and Technology*. https://asc.army.mil/web/news-hot-and-cold/

A Model for Evaluating the Maturity of a Modular Open Systems Approach

Alfred Schenker—works in the Software Engineering Institute's (SEI's) Software Solutions Division and has worked there for over 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 over 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]

Nickolas H. Guertin—was sworn in as Director, Operational Test and Evaluation on December 2021, then later in December 2023 as the Assistant Secretary of the Navy for Research, Development, Test, and Evaluation. A presidential appointee confirmed by the U.S. Senate, he serves as the Navy and Marine Corps Acquisition Executive, in charge of over $150 billion of product developments, fielding, and sustainment for the naval services.

Guertin has an extensive 4-decade combined military and civilian career in submarine operations; ship construction and maintenance; development and testing of weapons, sensors, and 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. [nhguertin@sei.cmu.edu]

Douglas Schmidt—is the Cornelius Vanderbilt Professor of Computer Science at Vanderbilt University. He is also a Visiting Scientist at the Software Engineering Institute at Carnegie Mellon University, where he served as the Deputy Director of Research and Chief Technology Officer from 2010 to 2012. He was recently confirmed as the Director of Operational Test and Evaluation, where he evaluates the effectiveness, suitability, survivability, and, when necessary, the lethality of systems produced by the U.S. Department of Defense.

Schmidt's research over the past 4 decades covers a range of software-related topics, including patterns, optimization techniques, and quality assurance of frameworks and model-driven engineering tools that facilitate the development of mission-critical middleware for distributed real-time embedded systems and intelligent mobile cloud computing applications running over wireless/wired networks and embedded system interconnects. [schmidt@dre.vanderbilt.edu]

Abstract

Defense acquisition leadership has long espoused the benefits of a Modular Open Systems Approach (MOSA). The discussion has been consistent, but the actions have not. We suggest that there is a spectrum of MOSA "compliant" implementations among projects. We refer to this as a spectrum of "MOSA maturity." The acquisition community would benefit from an evaluation framework—based on a model of MOSA maturity—to characterize how well MOSA-related policy objectives are being met. We suggest that a coherent set of attributes be investigated, and that results be assessed, to see if a program, system, system-of-systems, or enterprise has made the necessary changes to business, technical, and organizational models.

This paper describes an analysis construct that characterizes how well a weapon (or cyber–physical) system product has progressed in achieving the attributes of a MOSA. We will consider recently published attributes and criteria for MOSA as described by the Department of Defense (DoD), the Military Services, and Congress. We tie this work with emerging development practices to determine a more effective means of measuring and comparing MOSA capabilities across programs.

This approach, built on prior research (as well as its measures) aligns with the newest Military Services MOSA policies and the latest DoD Instruction 5000.02: Operation of the Adaptive Acquisition Framework. We identify new findings for the consistent application of MOSA practices in programs.

Keywords: Software Decomposition, Software Patterns, Integration, Modularity, Containerization, Virtualization, Acquisition, Business Model, Interoperability, System-of-Systems, Cyber-Physical Systems, Real-Time, Safety-Critical, Cyber-Secure, Intellectual Property, Modular Open Systems Approach, Open Systems Architecture, Maintainable, Durable, Self-Healing, Automated Testing, Configuration Management, Payloads, Platforms

Introduction

This paper introduces a way to characterize maturity with respect to implementing a Modular Open Systems Approach (MOSA) for Department of Defense (DoD) acquisition. We believe this characterization will lead to an improved evaluation method. We describe a hierarchy of business and technical acquisition aspects related to openness that is aligned to the most recent DoD acquisition policy instruction, DoD Instruction 5000.02: *Operation of the Adaptive Acquisition Framework* (i.e., Adaptive Acquisition Framework), and The William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021 (i.e., FY-21 NDAA; U.S. Congress, 2021). FY-21 NDAA Section 804 ("Implementation of Modular Open Systems Approaches") is provided along with a "hand-in-glove" separate section related to technical data rights. These sections together illuminate some specific requirements associated with MOSA for the DoD. We then connect those requirements with a tool that can be used to evaluate the cost of making investments in MOSA-aligned products.

This paper builds on recent work by Carnegie Mellon University's Software Engineering Institute (CMU/SEI) that evaluates open architecture approaches and other prior works on assessing acquisition approaches. Particularly noteworthy in this body of work is the blog post "Addressing Open Architecture in Software Cost Estimation," which deals with cost estimation in open architecture software-intensive systems (Gagliardi et al., 2020).

Background

Broad application of a MOSA across the DoD and Military Services of the Army, Air Force, Space Force, Navy, and Marine Corps (i.e., the Services) enables effective decision-making for the U.S. government in evaluating choices among innovative alternatives and competing technologies. A key motivation for a MOSA is to enable a mechanism for inserting innovative technical solutions from DoD providers (as robust and effective tools) into the hands of the military users (i.e., warfighters) as rapidly and affordably as possible.

At its core, however, a MOSA is an architectural constraint that should be balanced against other architectural constraints (such as performance, safety, and security). While principles of modularity and openness can be applied broadly, when it comes to a MOSA, the real benefit of MOSA occurs when the government correctly anticipates the specific pieces of technology that are likely to be upgraded/replaced over the life cycle and makes the necessary investments in that technology *when the product is being developed* to facilitate those changes/upgrades, thereby proactively reducing technical debt over the life cycle.

An effective MOSA should be implemented with (1) sound and mature technical characteristics, (2) well-reasoned and nuanced approaches to competitive dynamics, and (3) the thoughtful use of intellectual property rights in technical data. The key benefits of a MOSA-based implementation include the following:

- Enhance competition by employing open architectures with severable modules, allowing open competition of architectural functions/system components.
- Facilitate technology refresh by enabling delivery of new capabilities or replacement technology with minimal impact on system design.
- Incorporate innovation by ensuring operational flexibility to configure and reconfigure available assets to meet rapidly changing operational requirements.

- Enable cost savings/cost avoidance through reuse of technology, modules, or components from any qualified supplier across the acquisition life cycle.
- Improve interoperability by allowing changes and updates to severable software and hardware modules independently.

While it is hard to argue against the benefits of a MOSA from a technical or cost perspective, efforts to achieve them have been inconsistent at best and counterproductive at worst. A reliable and repeatable means of evaluating a MOSA would help guide MOSA implementations. Prior efforts to measure instantiations of a MOSA have had strengths and weaknesses, which we used to inform our approach described in this paper.

Congressional Direction

Congress has provided legislation in the FY-21 NDAA that documents a set of requirements for the DoD to achieve (U.S. Congress, 2021). In FY-21 NDAA Section 804 ("Implementation of Modular Open Systems Approaches"), a study of this language is instructional to parse the progression of a MOSA:

> *(a) Modular Open System Approach Requirement. — All major defense acquisition programs shall be designed and developed, to the maximum extent practicable, with a modular open system approach to enable incremental development and enhance competition, innovation, and interoperability. Other defense acquisition programs shall also be designed and developed, to the maximum extent practicable, with a modular open system approach to enable incremental development and enhance competition, innovation, and interoperability.*

This legislation has several elements that require an integrated business and technical strategy to achieve modularity, characteristics of the interfaces between those modules, the use of consensus-based standards to design those interfaces, and related acquisition requirements.

This legislation also provides detail with respect to system architecture:

> *(C) uses a system architecture that allows severable major system components and modular systems at the appropriate level to be incrementally added, removed, or replaced throughout the life cycle of a major system platform to afford opportunities for enhanced competition and innovation*.

It is especially noteworthy how this language has been modified from prior DoD instructions and guidance. The word *modified* has been removed from the list of characteristics that a modular approach should be able to provide. The legislation also provides a well-thought-out update to intellectual property rights in technical data that the government should employ for military systems. These rights include the ability to share information related to interfaces regardless of the nature of data rights associated with the underlying module. This change further informs us on how to characterize the maturation of MOSA in a program.

A succinct list of what these practices are expected to yield is also provided:

> *(i) significant cost savings or avoidance;*
>
> *(ii) schedule reduction;*
>
> *(iii) opportunities for technical upgrades;*
>
> *(iv) increased interoperability, including system of systems interoperability and mission integration; or*
>
> *(v) other benefits during the sustainment phase of a major weapon system.*

This list does not correlate directly to objective characteristics that translate to measurements and can be plugged into a formula to give a numerical result in terms of rating one program against another. However, both technical and business practices can be established that will guide a MOSA for the range of systems that Congress is interested in, including

- major system platforms
- major system components
- subsystems
- assemblies

Congress makes sure that interfaces are defined in a way that leads directly to business outcomes. The description it provides "goes to the heart" of modules that facilitate flexibility in composing new functions and outcomes that can be decoupled and connected in new ways across an array of military uses.

> *The term "modular system* interface" *means a shared boundary between major systems, major system components, or modular systems, defined by various physical, logical, and functional characteristics, such as electrical, mechanical, fluidic, optical, radio frequency, data, networking, or software elements.*
>
> *The term "modular system" refers to a weapon system or weapon system component that—*
>
> *(A) is able to execute without requiring coincident execution of other specific weapon systems or components;*
>
> *(B) can communicate across component boundaries and through interfaces; and*
>
> *(C) functions as a module that can be separated, recombined, and connected with other weapon systems or weapon system components in order to achieve various effects, missions, or capabilities.*

In defining MOSA, Congress established a set of verification criteria for the interfaces of these modular elements, enabling the government to measure something to ensure that products are meeting MOSA objectives. Specifically, it requires the following:

(i) comply with, if available and suitable, widely supported and consensus-based standards; or

(ii) (information related to the interfaces are delivered with rights to the technical data that allow sharing such that):

I. software-defined interface syntax and properties, specifically governing how values are validly passed and received between major subsystems and components, in machine-readable format;

II. a machine-readable definition of the relationship between the delivered interface and existing common standards or interfaces available in department interface repositories; and

III. documentation with functional descriptions of software-defined interfaces, conveying semantic meaning of interface elements, such as the function of a given interface field.

This section on data rights makes some distinct changes in rights to data associated with interfaces. It also clarifies a set of business practices that address the right to share information related to interfaces, regardless of the funding source.

In FY-21 NDAA, Section 1833 ("Proprietary Contractor Data and Rights in Technical Data") is decoupled from Section 2320, which is the MOSA section that makes substantive changes to the law regarding the government's rights in technical data (U.S. Congress, 2021). These sections provide much clearer statutory authority and identify a preference for rights in data. The union of an intellectual property strategy that is propelled by rights in technical data has long been a correlated practice in achieving MOSA objectives.

This Technical Data Rights section also facilitates sharing the inner designs of a product while seeking to sustain competitive dynamics in the limited defense market through the use of a little-used but long-established data right type called *Program Purpose Rights*. The context of this particular preference is to align a set of organizations around a collection of like products or technical domains where deep sharing and interaction are needed to field a complex and interoperable capability while preserving competitive pressures outside of those specific circumstances and in unrelated domains.

Taken in aggregate, the message in this legislation is clear:

- The technical architecture should be built on a set of standards that are open and available to any qualified provider.
- A modular construct for weapon systems must comport to business practices that facilitate the government's ability to choose alternatives in a competitive environment.
- Complete details of the interfaces that characterize the interaction between the modules must be made available to the government and can be provided to competitors in a related market.
- Modular designs and related interfaces will be subject to government verification and validation.
- Sharing information that represents the fire of innovation, which is the principal driver of a competitive market dynamic, must be preserved (Guertin & Womble, 2012).

Measures of MOSA implementation will need to address the requirements of this legislation.

DoD and Military Services MOSA Instructions and Guidance

In January 2019, the secretaries of the Military Services signed the memorandum for service acquisition executives and program executive officers (DoD, 2019) on the subject of *Modular Open Systems Approaches for our Weapon Systems in a Warfighting Imperative* (i.e., The Tri-Service MOSA Memo). In this seminal document, these secretaries not only identified an imperative, but they also provided specific examples of how to achieve it.

Here too, these leaders focused on standards for systems architecture and a need to drive data interoperability to "ensure our future weapon systems can communicate and share across domains." This directive provides grounding about mechanisms, which can be built on, to establish MOSA maturity measurements, including data interoperability.

Earlier Efforts at Measurement for MOSA

Modular Open Systems Approach, Program Assessment and Rating Tool

The MOSA Program Assessment and Rating Tool (PART) was an early effort by the Open Systems Joint Task Force (OSJTF, 2004), which operated from 1994 to 2004. The goal of the

MOSA PART was to characterize the degree to which the prior goals of the MOSA initiative were addressed. It identified the following five indicators:

- enabling environment
- modular design
- key interfaces
- open standards
- conformance

Per the OSJTF description,

> *MOSA PART is intended for use by DoD Program Managers as a means to assess their implementation of MOSA throughout the acquisition life-cycle. The MOSA PART is an analytic tool that evaluates responses to a set of interrelated questions to provide acquisition program executives with an objective and evidence-based assessment of the degree that MOSA is implemented in a program.*

The OSJTF established some valuable starting points for evolving the notion of an open system by identifying a distinction between open key interfaces correlated to the use of open standards.

Limitations. Participation in the OSJTF's MOSA PART was strictly voluntary. The MOSA PART was therefore unable to provide discernable metrics for the elements of each of these five measures in a way that could be used as criteria for a detailed assessment.

The Open Architecture Assessment Tool

The Naval Open Architecture Enterprise Team (OAET) used the MOSA PART as a starting point on which to develop the OA Assessment Model (OAAM), which is illustrated in **Figure** 1. This model developed two dimensions of program openness along the axes of business and technical openness (Open Architecture Enterprise Team [OAET], 2009).

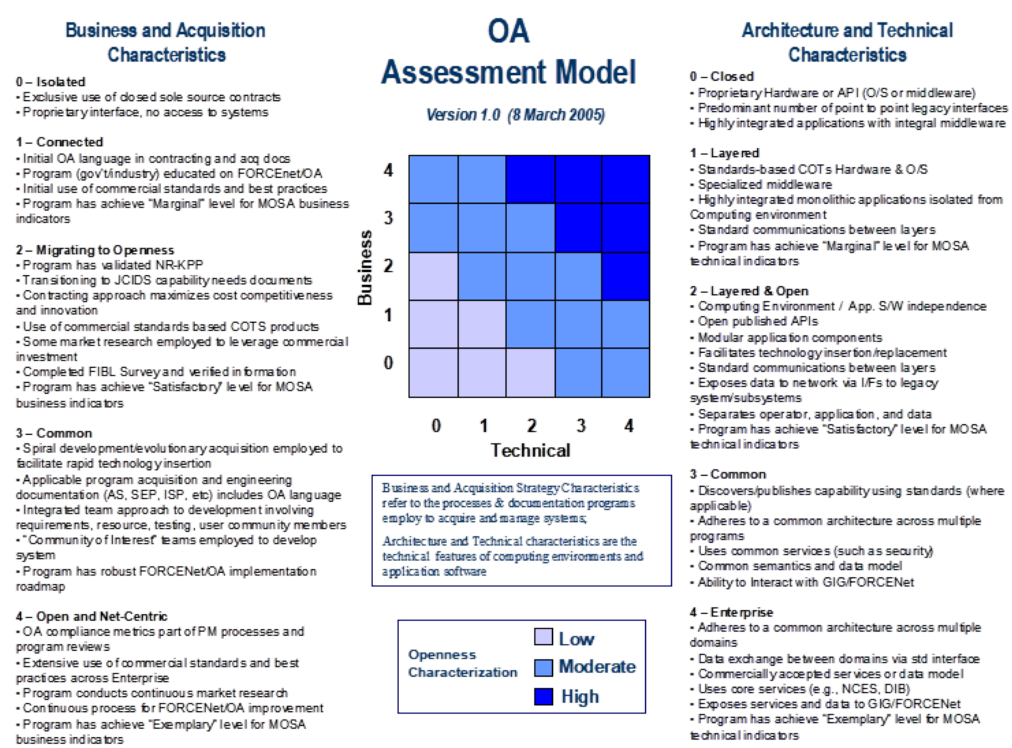


Figure 1. Open Architecture Assessment Model (OAET, 2009)

The Open Architecture Assessment Tool (OAAT) was developed by the OAET in response to leadership demands for some way to measure a degree of openness for a program. According to the *OAAT User's Guide*,

> *The OAAT is a tool for the use of Program Managers (PMs) and their OPNAV resource sponsors to assess, on a continuing basis, the OA maturity of a program and its systems. For complex programs, an assessment can be conducted on the whole program or individually on each significant sub-element of the program. In either case, the activity on which the assessment is conducted is called the unit of assessment. The OAAT assessments provide a current state of the unit of assessment that can then be used in conjunction with other factors, such as remaining service life, stage in the acquisition process, and potential for the system to change over time, to be compared with a desired state of openness.* (OAET, 2009)

The most current version of the OAAT is version 3.0, which was released in 2010. It has 64 questions that are roughly evenly split between business and technical characteristics. While most questions are optional, one-third of the assessment questions must be answered. These questions have a greater impact on the overall score as each one has a three-times multiplier for the scoring algorithm.

In addition, five of those questions were deemed so impactful that if the answers were not addressed above a threshold level, the overall score is capped at 50%. These litmus test questions eliminate the possibility of a program doing well in many small ways, while not addressing high-impact areas, yet achieving an artificially high numerical result. The OAAT is still available on the Defense Acquisition University's (n.d.) website and referred to in guidance documents used by both the DoD and the Services.

Limitations. The OAAT yields a single, two-dimensional numerical output based on the OAAM after taking in dozens of inputs stretched out over multiple technical, business, and cultural measures. As a tool, it does not provide insight on what measures have the greatest impact on the overall objectives of MOSA, nor does it provide a hierarchy of what measures are most important. Moreover, the OAAT provides a limited ability to compare implementation approaches across programs, thus providing little in the way that metrics can be used to guide cross-program or enterprise behavior.

While the tool was built to evaluate significant sub-elements (i.e., modules) as units of assessment, programs of record have not used the OAAT in this way. This lack of deeper analysis precluded developing any metrics on which to evaluate modular dimensions of openness. This lack of penetration in evaluation further eliminated an approach for characterizing the intersystem, or intra-program interfaces or interoperability performance.

Another limitation is that the business and technical objectives of Congress's requirements and DoD policy have matured significantly since 2010. Modularity and managed interfaces within and across systems and environments were not a factor in designing the OAAM and the subsequent OAAT.

Programs that were using the OAAT sought guidance from the naval OAET from 2007 to 2013. Each program completing the evaluation was satisfied with the score it received (regardless of outcome) and took no further action to improve its score. The OAET subsequently abandoned the OAAT as an input to the Navy's quarterly OA Report to Congress in favor of the *Naval Open Systems Architecture Questionnaire and Guidance* (i.e., NOA Questionnaire), which was designed to facilitate and advance an updated Naval OA Strategy and provide insight across programs and organizations (OAET, 2014; U.S. Navy, 2012).

NOA Questionnaire

In 2012, the Navy's acquisition executive changed the nature of that Service's plan to achieve an enterprise-level *Open Systems Architecture Strategy* (OAET, 2014). The naval OAET developed a related *Naval Open Systems Architecture Strategy* (U.S. Navy, 2012) that was released in that same year. As a part of that strategy, new measures provided a means by which cross-program comparisons could be made at a more detailed level than is facilitated in the OAAM. The NOA Questionnaire (shown in **Figure 2**) was developed as a limited set of questions that addressed the most impactful elements, many of which were extended from the MOSA PART and the OAAT (OAET, 2014).

Question	Yes/No
1. Have you developed an open competition acquisition strategy that enables contracting with third-party developers for modules that can be competed? Q1 Guidance	
2. Have you published a data rights/intellectual property strategy? Q2 Guidance	
3. Did you use the Naval OSA or DoD OSA Contract Guidebook for Program Managers to help contract or RFP development? Q3 Guidance	
4. Did your current contract or RFP call for government non-development items and COTS (including open source or community source components)? Q4 Guidance	
5. Did your current contract or RFP disclose the technical architecture and supporting systems Engineering information adequately to enable third-party developers to participate? Q5 Guidance	
6. Did your current contract or RFP call for replaceable/refreshable components that can be re- competed? Q6 Guidance	
7. Does your program acquisition plan call for re-competition of the system and/or components every 3-7 years? Q7 Guidance	
8. Do you plan to issue any RFPs in the next 3 fiscal years? Q8 Guidance	
9. Have you measured the "openness" of your systems (e.g., used the OAAT)? Q9 Guidance	
10. Have you identified potential modules that can be competed? (e.g., that can be replaced, modified, upgraded, or extended) Q10 Guidance	
11. Do you employ a modular, open systems approach (MOSA)? The use of open standards for key interfaces is part of this approach. Q11 Guidance	
12. Did you use open standards-based designs and agreed upon data models that are being used by a different program or are being facilitated by a COI? Q12 Guidance	
13. Have you formally documented the technical framework of your system? Q13 Guidance	
14. Have you developed an asset reuse strategy in accordance with Naval Enterprise Policy? Q14 Guidance	
15. Is your contractor following your Open Systems Management Plan (OSMP)? Q15 Guidance	
16. Does your acquisition strategy or life cycle support plan include periodic technology refresh or capability insertion approaches for life cycle affordability? Q16 Guidance	
17. Have you implemented a training program to educate your acquisition work force on OSA (e.g., continuous learning modules, workshops)? Q17 Guidance	

Figure 2. The NOA Questionnaire (U.S. Navy, 2012)

The 17 questions (eight business, eight technical, and one workforce) came with guidance and information needed to understand what characteristics of a program would yield a positive response.

The results were reported to Congress quarterly for the following 2 years, and these results facilitated cross-organization and cross-program measurements of progress towards achieving the overall objectives of the *Naval Open System Architecture Strategy*.

The NOA Questionnaire results were used by both program managers and their associated program executive offices (PEOs) to understand how to improve overall Open Systems Architecture approaches and achieve the objectives of the *Naval Open Systems Architecture Strategy*.

Limitations. The NOA Questionnaire is built on yes/no responses and was developed to drive reporting and cross-program/organization behavior, not to directly assess details of implementations.

The primary purpose of the questionnaire was to support the *Naval Open System Architecture Strategy* and the secretary of the Navy's quarterly report to Congress to compare progress across programs. The *Naval Open System Architecture Strategy* and a need to perform these surveys and subsequent analyses were not codified into long-term policy. After these reports were no longer required, the drive to execute an enterprise strategy faded, and the need to participate in the questionnaire was truncated.

Office of the Under Secretary of Defense for Research and Evaluation MOSA Assessment Criteria

In 2022, as part of the response to the FY-21 NDAA, the Office of the Under Secretary of Defense for Research and Evaluation (OUSD[R&E], 2022) released criteria for assessing a MOSA. The DoD had established a Modular Open Systems Working Group (MOSWG), and in 2018 the MOSWG stood up an Assessment Tiger Team to survey the use of MOSA in DoD acquisition programs. In 2021, the Tiger Team reported that “although it had identified general criteria for assessing the effectiveness of MOSA compliance, it had not agreed on specific criteria that would be applicable across all Service and program types.”

The DoD had previously defined a set of MOSA tenets, referred to as “pillars,” to guide the use of MOSA in defense acquisition programs. These pillars, defined in 2011, are remarkably like the indicators defined in the previously described PART, although they have been elaborated in much more detail.

The MOSWG decided to require the Services to explicitly connect their tailored assessment criteria to these pillars. An example of how to do that was provided in the assessment guidance document.

Limitations. The guidance document and the assessment criteria were produced relatively recently and have not had the opportunity to be put into practice.

Open System Verification Demonstration

As part of the acquisition plan for the U.S. Army’s Future Attack and Reconnaissance Aircraft (FARA), the Army planned a series of open system verification demonstration (OSVD) events to assess the degree to which the FARA contractor’s designs met the Army’s MOSA standards (Sikorsky, 2023). The Army had provided a set of MOSA scenarios to the contractors as part of the acquisition government-furnished information (GFI). The demonstration was to verify the government could replace a major system component with the following constraints: (1) by using nothing but the contractor’s TDP, (2) using an independent third party to implement the component replacement, and (3) performing the work in the contractor’s Systems Integration Lab (SIL).

It was expected that there may be training and orientation required to ensure the independent third party is fluent with the contractor’s development environment, so initial demonstrations were focused on learning how to make the change (as opposed to a complicated component replacement) and were therefore relatively simple. When they had demonstrated competency with the development environment, the third party moved on to more challenging component replacements.

This approach resulted in a much more involved demonstration than what had been done in prior MOSA assessment methods. The results, when made available, should make an interesting read. This type of assessment requires a level of financial commitment (by the U.S. government) to perform the component replacement and demonstrates the importance that the government has placed on achieving a MOSA. Unfortunately, the FARA program was cancelled by the Army in early 2024, so we won’t know what the outcome would have been, but clearly the approach is worthy of a mention in a paper on this topic. It was expected that the results of the assessment would have influenced the selection of the winning contractor (as part of the source selection process).

Limitations. The OSVD assessment represents one of the first times that the government has tried to assess openness and modularity in such a tangible way. Prior attempts often focused on design or architecture documentation and review, falling short of actually replacing a major system component. We believe this type of assessment provides direction to becoming the “gold standard” for MOSA assessment, but more experience with performing it is needed. For example, the opportunity to collect data (e.g., effort, issues, lessons learned) regarding the experience of making the change is unparalleled. A standard set of measures must be developed to support this type of assessment.

MOSA Maturity

MOSA does not happen by accident. It requires a deliberate effort by an organization to accomplish specific objectives for their products. The broad benefits of MOSA have been described above, but how do we know that (1) our organization possesses the knowledge and skills needed to develop a strategy to acquire products following MOSA principles, and (2) our source selection process will produce a contractor that correctly applies the MOSA principles to the design and integration of our products? A MOSA Maturity model could be used to help define and assess the competencies of both the acquirer and the contractor and could incorporate the pro forma approaches that have been attempted over the past 20 years.

Is it simply adequate for a project to satisfy the measurement criteria of a particular assessment? Or are there other indicators of an organization’s experience with MOSA that would provide more insight for an organization?

For a contractor hired to develop a “MOSA compliant” product, we believe that there is a spectrum of MOSA compliance that ranges from “box-checking” to “the way we do things around here.” We believe that there are qualitative indicators, including such items as

- how models are used (e.g., data models, MBSE) in the design
- how the interfaces are documented
- how much due diligence was spent on MOSA (i.e., effort spent performing trade-off analyses where MOSA was one of the criteria)
- experience with product lines and product line governance

There is also an expected level of experience needed on the acquirer side. In fact, a critical element of acquirer competency is to be able to discern the differences between contractors competing on contracts that require a MOSA. Additionally, there may be competency needed in the following areas:

- experience with product lines and product line governance
- elaborated scenarios (or use cases) that illustrate the intent of the MOSA
- existence of data models that are used in the product domain
- experience with model-based methods for specifying requirements
- standard measures for how to characterize the MOSA implementation

Maturity of Data and Interfaces

The Tri-Service MOSA Memo and the FY-21 NDAA make clear that interoperability is based on the interfaces between major elements, the standards on which those interfaces are built, and the intelligible structure to the data so that the products can be mixed and matched across a diverse set of military capabilities.

As the use of a module (be it in a system, a platform, or a product) is expanded to other areas, portability and multi-context interoperability are predicated on the ability to consume and provide information in other arenas or domains. Interface documentation, including clarity of semantics and syntactics, is then critical to achieving the objectives of a MOSA strategy.

The Interface Documentation Maturity Levels (IDML) model, shown in **Figure** 3, was developed to establish a progression of characteristics needed to address how to develop interfaces that support a MOSA strategy (Hand et al., 2018).

Figure 3. Interface Documentation Maturity Levels

This approach to establishing an interface maturity construct illustrates how to create a MOSA maturity model.

Open Systems Architecture Configurability Rating Checklist Tool

Figure 4 shows the structure of a tool that the Software Engineering Institute (SEI) developed that enables a program to perform an Open Systems Architecture assessment on a selected software architecture and then provide software cost-estimation inputs, including assessment ratings, to a standard software cost-estimation program. This tool goes into greater depth of the characteristics of MOSA; it examines a product through the lens of a separate units of assessment and addresses the following:

- **Modularity**: System architecture key components are encapsulated, cohesive, self-contained, and loosely coupled.
- **Interface Standards**: A widely available document exists that specifies interfaces, including services provided/required, protocols, message and data formats, and so forth.
- **Layering and Tiers**: A software abstraction provides separation from other software packages and technology.
- **Open and Accessible Standards**: Key interfaces are based on open and accessible standards that are widely used, consensus-based, published, and maintained by recognized communities of interest.

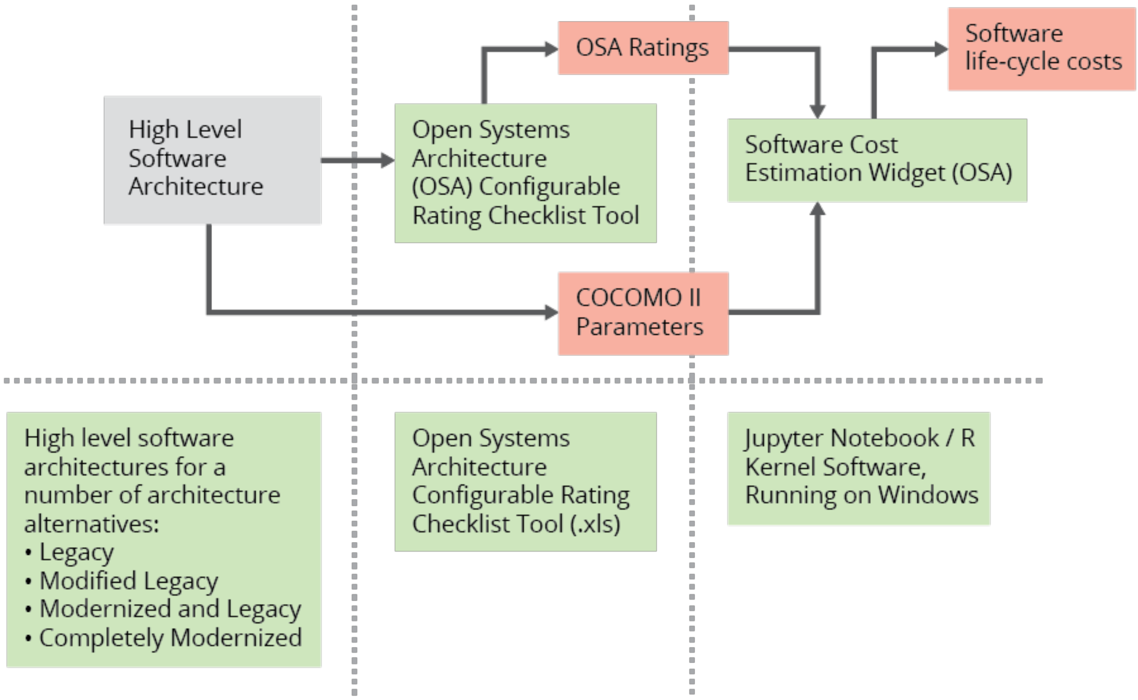


Figure 4. Open Systems Architecture Configurability Rating Checklist Tool

As an example application of this tool, a representative acquisition program was evaluated to assess the cost performance of keeping a legacy design against making an up-front investment to open the program and improve the overall architecture to facilitate improved reliability, maintainability, and upgradability. **Figure 5** shows the analysis results.

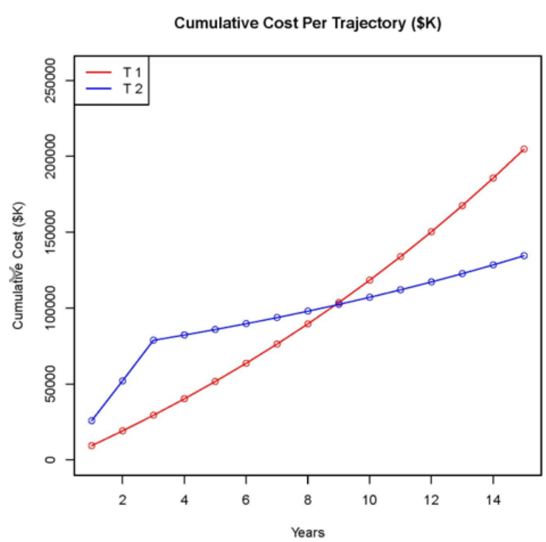


Figure 5. Cost Assessment of Adopting MOSA Versus Staying the Course Using the Open Systems Architecture Configurability Rating Checklist Tool

The goal of this work is to remove uncertainty about cost as a barrier to adopting open systems architecture methods, platforms, and tools. However, this tool is limited by the quality of the input data, which is based partially on the OAAT. To improve this tool and broaden its applicability, a more up-to-date assessment of MOSA maturity is needed.

MOSA Maturity Model

Informed by these past efforts, we blended the legislative requirements from Congress with the acquisition policy needs of the DoD to create a hierarchy composed of criteria that address the business needs and technical discipline MOSA requires for a product, system, or platform. We continue to assert the need to evaluate the framework of the technical architecture to be as important as the management of the acquisition approach to achieve the objectives of MOSA.

We end this paper with "the model," which is instantiated as a set of scenarios, broken into three tiers, ranked by importance, and split along the dimensions of business and technical characteristics. We can use these scenarios to ascertain how well the MOSA goals are being met, which can be assessed through evidence-based measures and logic tests.

	Business	Technical
Growing	Can a new module be added to a product to improve its fielded performance (i.e., innovation) within a week of completing integration testing?	Does the interface of the module have well-defined and published semantics and syntactics (i.e., data model) for interoperability that are addressable by any other defense program?
	Is the technical architecture for the current design documented in a digital model and made available to any qualified party?	Is there sufficient documentation or a digital model so that the role of the system integrator can be competed or subsumed by the government with minimal effort?
		Is there sufficient documentation or a digital model for a module so that the role of the product provider can be competed or subsumed by the government with minimal effort?
Mature	Is the module's performance documented in a digital model that can be used for the competition of existing capabilities?	Can a different module replace an existing module within a day with the same or fewer integration errors?
	Is there an intellectual property strategy that has been validated against the newest data rights legislation, including a preference for Program Purpose Rights?	Can modules be upgraded or replaced quickly either directly or by technicians in the field?
	Are the interfaces of the module, system, or platform published (either in a digital model or in a document) and made available to any qualified organization?	Is the software environment made up of an open platform (e.g., containerization construct or micro-service architecture) that is widely published or available to any qualified competitor?
Compliant	Can an existing module (e.g., major system component) be integrated into a different domain within a month of a new domain being identified?	Is a module sufficiently decoupled from an interface standard so that it can be repurposed or upgraded to use a different interaction mechanism?
	Is there an open competition acquisition strategy that enables nonincumbents to compete and win as alternative providers?	Can an existing module be upgraded to operate in a new environment or a different warfighting domain within 3 months of that new domain being identified?
	How often is the incumbent's implementation of an Open System Management Plan validated by an independent third party?	Are the modules sufficiently decoupled from their execution platform so that an update to hardware or other infrastructure can be performed in a week?
	Can a module be incrementally changed and deployed with known effects to other modules it interacts with?	Can a module be replaced with an alternative either for programmatic reasons or improved performance?
Progressing		Can the module execute without coincident execution of other specific weapon systems or components?

	Business	Technical
	Can an existing module (e.g., component in a major system platform) be added, removed, or replaced throughout the life cycle?	If the module has sensitive timing needs, is there a validated model of the interaction with other related modules that others can use to evaluate replacement alternatives?
Early	How often are the members of the systems, development, and operations teams provided with training on the implementation of a MOSA?	How often are the members of the systems, development, and operations teams provided with training on the implementation of a MOSA?
	Can modules of a system or platform be severed from its original deployment for use in other contexts?	Does the module construct exist across implementation domains of electrical, mechanical, fluidic, optical, radio frequency, data, networking, or software elements?
None	How many modules of the system will be competed in the next 3 to 7 years?	Can a product roll back to an older safe state if a replacement becomes unstable or inoperable?

Why This Approach Is Different

MOSA is an evolving practice in both depth and breadth. It has changed since the early days of the OSJTF and other hallmark programs that informed an open architecture approach for the DoD (Guertin & Miller, 1998). The details matter, and measures that address needed change can inform progress. Using a scenario-based approach facilitates the evolution of the methods, while the characteristics of what is to be achieved remains somewhat stable. Any product, system, or platform can be evaluated by starting with basic levels and elevating the characteristics of what constitute both the technical and business steps to achieving the goals of a MOSA.

Road to Adoption. The following activities should be put into place to facilitate a global set of MOSA maturity measurements that inform leadership and elevate best practices for all programs:

- Validate these proposed measures against selected products, programs, and platforms to baseline the nature of MOSA maturity. Have those measures independently verified.
- Use that baseline to inform changes to the measures prior to full deployment to all programs.
- Capture those validated measures as inputs to the DoD and Services.
- Develop and deploy a set of matching DoD and Services policies that require all programs of record, including programs that operate under larger acquisition category arrangements, to perform the new assessment. Have a third party validate the responses.
- Perform a data analysis to identify needed next steps and evaluate efforts that best meet the spirit and the letter of the law and policy.
- Report the findings to Congress to show progress against its requirements.

Barriers to Adoption. If there is not a requirement for assessing all DoD programs with respect to their implementation of MOSA, only those who expect to get a great score will perform the assessment, and enterprise value will not be achieved.

Performing independent validation is a lesson learned from the limited utility of the results from the OAAT and MOSA PART. However, independent validation requires a cadre of competent MOSA validators. Other maturity models (e.g., CMMI) struggled with qualification of the independent validators and, depending on how the validator was contracted (by the government or by the contractor), maintaining their independence. Inconsistently implemented approaches within the Military Services and across the DoD will limit the ability of achieving a whole-of-government comparison and identification of enterprise value to improve overall robustness and transparency.

Summary

Measurements of MOSA have existed for a long time and have their share of weaknesses. The OAAT builds on the MOSA PART, but the results of a single measure of technical versus programmatic openness is too coarse to provide effective assessment of adherence to MOSA principles and requirements. The NOA Questionnaire is not detailed enough in assessing critical aspects of a program to capture specific measures that can be addressed to improve outcomes, though it does provide a mechanism that facilitates cross-program and cross-organization comparison. The SEI's Open Systems Architecture Configurability Rating Checklist Tool is informative to acquisition managers looking to make a set of clear business and technical choices, but it should be informed by measures that comport to the current requirements of Congress, the DoD, and the Services.

The next step is to take advantage of the lessons learned from these earlier MOSA-based measurement efforts and propel a new set of decisions based on sound technical and business measures that will also be flexible in addressing the evolving implementation methods. The methods applied to develop complex cyber–physical systems are always in motion, and any measurement strategy needs to account for this motion. The approach we use to account for these constantly changing methods is to keep the measures focused on outcomes.

Codifying an approach for measuring MOSA maturity and providing that as an input to new tools, such as the Open Systems Architecture Configurability Rating Checklist Tool, will support informed decisions at the module, system, and platform levels to improve warfighter outcomes.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM24-0386

References

Defense Acquisition University. (n.d.). *Modular open systems approach community of practice*. Retrieved March 26, 2024, from https://www.dau.edu/cop/mosa

DoD. (2019). *Modular open systems approaches for our weapon systems in a warfighting imperative*. https://www.dsp.dla.mil/Portals/26/Documents/PolicyAndGuidance/Memo-Modular_Open_Systems_Approach.pdf?ver=2019-01-18-122921-933

Gagliardi, M., Konrad, M., & Schmidt, D. (2020, July 6). Addressing open architecture in software cost estimation. *SEI Blog*. https://insights.sei.cmu.edu/blog/addressing-open-architecture-software-cost-estimation/

Guertin, N., & Miller, R. (1998). A-RCI—The right way to submarine superiority. *Naval Engineer's Journal*, *110*(2), 21–33. https://www.ingentaconnect.com/contentone/asne/nej/1998/00000110/00000002/art00010

Guertin, N., & Womble, B. (2012). Competition and the DoD marketplace. *Annual Acquisition Research Symposium Proceedings & Presentations*. https://dair.nps.edu/bitstream/123456789/1342/1/SYM-AM-12-076.pdf

Hand, S., Lombardi, D., Hunt, G., & Allport, C. (2018). *Interface Documentation Maturity Levels (IDML): An introduction*. Skayl. https://www.skayl.com/post/interface-documentation-maturity-levels-idml-an-introduction-1

Open Architecture Enterprise Team. (2009). *Open architecture assessment tool 3.0 user's guide.* U.S. Navy. https://www.dau.edu/cop/mosa/documents/oaat-v3

Open Architecture Enterprise Team. (2014). *Naval open systems architecture questionnaire and guidance.* U.S. Navy.

Open Systems Joint Task Force. (2004). *Program manager's guide to a modular open systems approach (MOSA) to acquisition, version 2.0*. U.S. Department of Defense. https://www.acqnotes.com/Attachments/Program%20Managers%20Guide%20to%20Open%20Systems,%20Sept%202004.pdf

Office of the Under Secretary of Defense for Research and Engineering. (2022). *MOSA assessment criteria*. https://www.cto.mil/wp-content/uploads/2023/06/MOSA-Assess-2022.pdf

Sikorsky. (2023). *RAIDER X digital backbone drives MOSA success for the Army.* DefenseNews. https://www.defensenews.com/native/Sikorsky-Lockheed-Martin/2023/07/10/raider-x-digital-backbone-drives-mosa-success-for-the-army/

U.S. Congress. (2021). *The William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021* (Pub. L. 116-283). U.S. Government Publishing Office. https://www.congress.gov/116/plaws/publ283/PLAW-116publ283.pdf

U.S. Navy. (2012). *Naval open systems architecture strategy*. https://www.researchgate.net/publication/319042060_Naval_Open_Systems_Architecture_Strategy

PANEL 23. ARTIFICIAL INTELLIGENCE ACROSS THE ACQUISITION LIFECYCLE

Thursday, May 9, 2024	
3:45 p.m. – 5:00 p.m.	**Chair: Randy Pugh, Colonel, USMC (Ret.),** Director, Naval Warfare Studies Institute, Naval Postgraduate School ***A Semiautomated Framework Leveraging NLP for Skill Identification and Talent Management of the Acquisition Workforce in the Department of Defense*** Jose E Ramirez-Marquez, Stevens ***Planning for AI Sustainment: A Methodology for Maintenance and Cost Management*** Iain Cruickshank, USMA/ACI ***System Acquisition Cost Modeling Initiative to Quantify AI Assistance*** Raymond Madachy, Naval Postgraduate School

Colonel Randy Pugh, USMC (Ret.)—joined the Naval Postgraduate School in 2019, where he served as the Marine Corps Senior Service Representative and Military Associate Dean of Research. He became Deputy Director of the Naval Warfare Studies Institute in 2020 and took on the role of Acting Director in 2022. In August 2023, he become the first permanent director of NWSI. In this position, he helps connect NPS to the Fleet Marine Forces and Headquarters Marine Corps and Navy on research topics of the highest priority and helps ensure that NPS' educational offerings satisfy the Navy's knowledge and skills requirements.

Colonel Pugh has spent the majority of his career as a Signals Intelligence / Electronic Warfare Officer, serving in command and staff billets at 1st Radio Battalion, as the SIGINT/EW Project Lead at Marine Corps Systems Command, as the Operations and Executive Officer at 3d Radio Battalion, and as the Commanding Officer of 2d Radio Battalion. He has deployed with the 31st Marine Expeditionary Unit (Special Operations Capable), I MEF, II MEF, Special Operations Command Pacific and Special Operations Command Europe to locations including Iraq, Afghanistan, and the southern Philippines. He recently served as the Commanding Officer of Marine Corps Intelligence Schools.

Colonel Pugh has a Master's degree in National Security and Strategic Studies from the Naval War College, a Master in Military Studies from the Marine Corps Command and Staff College, a Master of Science degree in Computer Science (Software Engineering) from Naval Postgraduate School, and a Bachelor of Science degree from United States Naval Academy. His NPS Master's thesis explored the use of artificial intelligence as a means to accelerate system integration.

A Semiautomated Framework Leveraging NLP for Skill Identification and Talent Management of the Acquisition Workforce in the Department of Defense

Dr. Jose E. Ramirez-Marquez—is an Associate Professor at the School of Systems and Enterprises, Stevens Institute of Technology. He holds degrees from Rutgers University in industrial engineering (PhD and MSc) and statistics (MSc) and from Universidad Nacional Autonoma de Mexico in actuarial science. His research focuses on developing mathematical models to analyze and compute system operational effectiveness - reliability and vulnerability as the basis for designing system resilience. He has published more than 200 refereed manuscripts related to these areas in technical journals, book chapters, conference proceedings, and industry reports.

Garry Shafovaloff—is the Senior Advisor, Policy and Legislation, Defense Acquisition University. Previously, he served as the Director and Deputy Director of the Office of Human Capital Initiatives (HCI), responsible for defense acquisition workforce strategic planning and initiatives from 2010 through 2022. In 2022, he served as a senior lead for the $50 million Artificial Intelligence DoD Upskilling initiative in partnership with the Chief, Digital and Artificial Intelligence Office. He is currently detailed to the Office of the Assistant Secretary of Defense (Acquisition), serving as the Senior Program Manager for the Defense Civilian Training Corps (DCTC) pilot initiative.

Mark Krzysko—is the Principal Deputy Director for Acquisition Policy and Innovation directing acquisition data governance, data access, and data science enabling the Department of Defense sound business decision-making. Additionally, he served in the National Academies of Sciences, Engineering, and Medicine's Data Science Post-Secondary Education Roundtable discussion on data science education and practice, the needs of the community and employers, and ways to move forward. 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.

Dr. Dinesh Verma—received a PhD (1994) and an MS (1991) in industrial and systems engineering from Virginia Tech. Verma currently serves as the Executive Director of the Systems Engineering Research Center, a U.S. Department of Defense–sponsored University Affiliated Research Center focused on systems engineering research, along with the Acquisition Innovation Research Center. At Stevens, he has proposed research and academic programs exceeding $175 million. He has authored over 100 technical papers, technical monographs, and three textbooks. Verma has received three patents in the areas of life-cycle costing and fuzzy logic techniques for evaluating design concepts.

Abstract

The Department of Defense (DoD) must address critical questions about talent management and workforce adaptability. This research introduces the potential for leveraging Natural Language Processing (NLP) techniques to address these challenges. The paper describes an NLP-based framework to analyze vast text data, including government, industry, and academic reports. The primary objective is to identify critical skills necessary within the DoD acquisition workforce efficiently and accurately. By automating this process, the DoD can swiftly pinpoint areas of expertise and allocate resources accordingly, ensuring the hiring and deploying of personnel with the right skills where needed most. With the insights derived from NLP analysis, decision-makers within the DoD can make informed choices regarding talent acquisition, training and development programs, and skill gap remediation. The ability to swiftly and accurately identify essential skills optimizes resource allocation, reduces skill gaps, and elevates operational efficiency. This newfound efficiency extends to talent management, enabling the DoD to nurture and develop critical skills proactively. Identifying and managing critical skills is pivotal for ensuring preparedness and resilience in a rapidly changing world order.

Keywords: Skills, Data Science, Natural Language Processing, Data Visualization

Introduction

The Department of Defense (DoD) stands as one of the most expansive and intricately structured organizations globally, charged with safeguarding the national security interests of the United States. In an era marked by evolving geopolitical tensions, particularly concerning the escalating influence of Russia and China, the DoD's capacity to effectively harness and optimize its human resources is paramount in maintaining a competitive advantage. Talent management transcends conventional recruitment and retention concepts; it encompasses identifying, cultivating, and deploying critical skills and expertise vital to addressing contemporary security challenges. As adversaries continually enhance their military capabilities and extend their influence, the agility and adaptability of the DoD hinge significantly on its ability to leverage advanced technologies, such as Artificial Intelligence (AI) and Natural Language Processing (NLP), to discern and cultivate the requisite skills necessary to outpace adversaries. Consequently, talent management strategies integrating cutting-edge NLP techniques can prove instrumental in ensuring the DoD's agility, responsiveness, and readiness in navigating complex and dynamic global threats.

The effective management of a workforce as expansive and multifaceted as the DoD is essential for upholding the nation's military readiness and global influence. By harnessing NLP and other advanced technologies, the DoD can streamline skill identification, align human resources with strategic objectives, and empower decision-makers to make well-informed decisions regarding recruitment, training, and skill enhancement initiatives. Within this framework, talent management emerges as a potent force multiplier, enabling the DoD to adeptly confront the nuanced and evolving challenges Russia and China pose. Ultimately, the efficacy of talent management strategies within the DoD significantly contributes to the United States' ability to assert itself globally and navigate the intricate dynamics of the contemporary international security landscape.

Problem Description

Amidst escalating geopolitical tensions, particularly with nations like Russia and China, the DoD grapples with myriad challenges in effectively identifying crucial skills and managing its workforce. The evolving landscape of modern warfare and rapid technological advancements necessitate continually adapting the DoD's workforce to anticipate and counter emerging threats. However, identifying these critical skills is complex due to the DoD's diverse composition, encompassing various military services, civilian roles, and contracted personnel. Geopolitical tensions introduce unpredictable dynamics, demanding a nimble workforce capable of addressing traditional military challenges alongside emerging threats like cyber warfare, information warfare, and hybrid conflicts. Balancing long-term skill development with the imperative for immediate readiness in a dynamically changing global environment further compounds this challenge.

Effectively managing the DoD's workforce in such circumstances demands a nuanced approach, considering demographic shifts, technological progress, and geopolitical realities. Furthermore, in an era where recruitment and retention strategies extend beyond talent attraction to include talent retention amidst heightened competition, the DoD faces aligning its strategies with its mission. Addressing these challenges is a matter of organizational efficiency and a critical component of national security, enhancing deterrence capabilities and ensuring a credible defense posture. Thus, amidst geopolitical tensions, the DoD's ability to identify and manage essential skills is pivotal in bolstering the United States' preparedness and resilience in an ever-evolving global context.

This research developed a framework that implements AI and NLP techniques to identify critical skills within the DoD workforce. By harnessing the capabilities of NLP, the project endeavors to enhance talent management, workforce planning, and skill development strategies within the DoD, ultimately contributing to a more agile and effective defense organization. The project's overarching objectives are as follows:

1. Skill Identification: Utilize NLP algorithms to analyze extensive textual data, including industry, government, and academic reports, to identify critical skills within the DoD workforce automatically.
2. Decision Support: Provide DoD decision-makers with actionable insights and recommendations from NLP analysis, empowering them to make well-informed decisions regarding talent acquisition, training programs, and skill gap remediation.

Importance

The proposed framework uses NLP techniques to identify pivotal skills within the DoD workforce to enhance defense operations, efficiency, and readiness. Against the backdrop of escalating geopolitical tensions demanding swift response and adaptability, the expeditious and accurate identification of essential skills holds immense importance in bolstering defense operations. By automating skill identification via NLP, the DoD can promptly identify areas of expertise and allocate resources accordingly, ensuring optimal deployment of personnel with the requisite skills to areas of utmost need. This streamlined process enhances resource allocation efficiency and augments readiness by mitigating skill gaps and elevating overall force preparedness.

The efficiency gains derived from NLP-driven skill identification extend beyond defense operations to encompass talent management and career development within the DoD. The research fosters strategic talent management initiatives, empowering the DoD to proactively nurture and cultivate critical skills and enabling targeted training programs and skill enhancement strategies. Ultimately, this fosters individual personnel effectiveness and contributes to cultivating a more agile and responsive defense organization adept at confronting the evolving challenges posed by geopolitical tensions. The ripple effect of the project's impact extends across various facets of defense, culminating in heightened operational efficiency and enhanced readiness, both indispensable attributes in navigating the intricacies of a dynamic and multifaceted global security landscape.

Literature Review

Automated Talent Management

NLP constitutes a branch of AI dedicated to endowing computers with the ability to comprehend, interpret, and generate human language meaningfully. Within talent management and human resources, NLP emerges as a transformative technology, presenting innovative solutions to enduring challenges and encompassing a spectrum of HR functions, from recruitment and talent acquisition to employee engagement and development.

One prominent application of NLP in talent management involves the automation of job descriptions and candidate resume analysis during recruitment processes. NLP-driven tools adeptly sift through job requirements and applicant qualifications, enabling HR professionals to swiftly pinpoint the most suitable candidates for specific roles (Vanetik & Kogan, 2023). Moreover, NLP-enabled sentiment analysis of job postings and social media activity offers invaluable insights into employer branding and aids organizations in gauging their perception of the job market (Allioui & Mourdi, 2023).

Furthermore, NLP plays a pivotal role in fostering employee engagement and retention. Through analyzing employee feedback, encompassing survey responses, performance evaluations, and informal communication channels like emails and chat logs, NLP identifies patterns and sentiment trends indicative of potential areas of concern or dissatisfaction. Timely detection of employee disengagement empowers HR teams to intervene proactively, enhancing retention rates and bolstering workplace satisfaction (Gomathi et al., 2023). Additionally, NLP facilitates the development of personalized learning and growth strategies by identifying individual skill gaps and recommending pertinent training resources.

NLP promises to revolutionize talent management and HR practices by automating mundane tasks, providing insights into employee sentiment, and facilitating data-driven decision-making. From streamlining recruitment procedures to enhancing employee engagement and development initiatives, the multifaceted applications of NLP contribute to realizing more efficient and strategic HR operations.

Skill Identification

As previously discussed, the process of skill identification entails analyzing extensive volumes of textual data to pinpoint specific skills possessed by individuals or required for particular roles. In this regard, NLP algorithms demonstrate exceptional efficacy, leveraging techniques such as Named Entity Recognition (NER) to extract skill-related keywords and phrases from unstructured text automatically. These algorithms adeptly discern subtle variations of skills, including synonyms or related terms, ensuring a comprehensive comprehension of an individual's or role's skill repertoire. Furthermore, NLP can contextualize these skills, distinguishing between incidental mentions and those integral to an individual's proficiency or role requirements.

Following skill identification, NLP offers the potential to streamline skill mapping processes by establishing connections between identified skills and specific job roles, career trajectories, or developmental pathways (Mohanty et al., 2023). Through analyzing skill-role relationships within extensive datasets, NLP algorithms uncover patterns and correlations, thereby generating skill-to-role mappings that are both data-driven and adaptable. This automated approach enhances talent management by furnishing decision-makers with precise insights into skill requirements across diverse roles and career trajectories. Additionally, NLP-powered decision support systems proffer actionable recommendations from skill analyses, empowering HR professionals and organizational leaders to make informed decisions about talent acquisition, training initiatives, and skill gap remediation. For instance, by scrutinizing skill data, NLP can propose tailored learning trajectories for employees, enabling them to cultivate critical skills aligned with their career aspirations and organizational imperatives (Caratozzolo et al., 2023). Overall, NLP's automation of talent management processes enhances operational efficiency, mitigates bias, and facilitates data-driven decision-making within the complexities of contemporary workforce environments.

Framework

Efficiently handling extensive textual data sets poses a significant contemporary challenge in information management. This section introduces a novel framework designed to streamline content extraction, text summarization, and the creation of executive summaries. These tasks hold critical importance across diverse domains, ranging from academic research to corporate decision-making, facilitating rapid and informed information retrieval and decision support. As illustrated in Figure 1, the framework comprises two key phases: Phase 1 encompasses Text Extraction and Summarization, while Phase 2 focuses on Skill Identification and Analytics. Notably, all framework processes have been developed utilizing AI-assisted technologies.

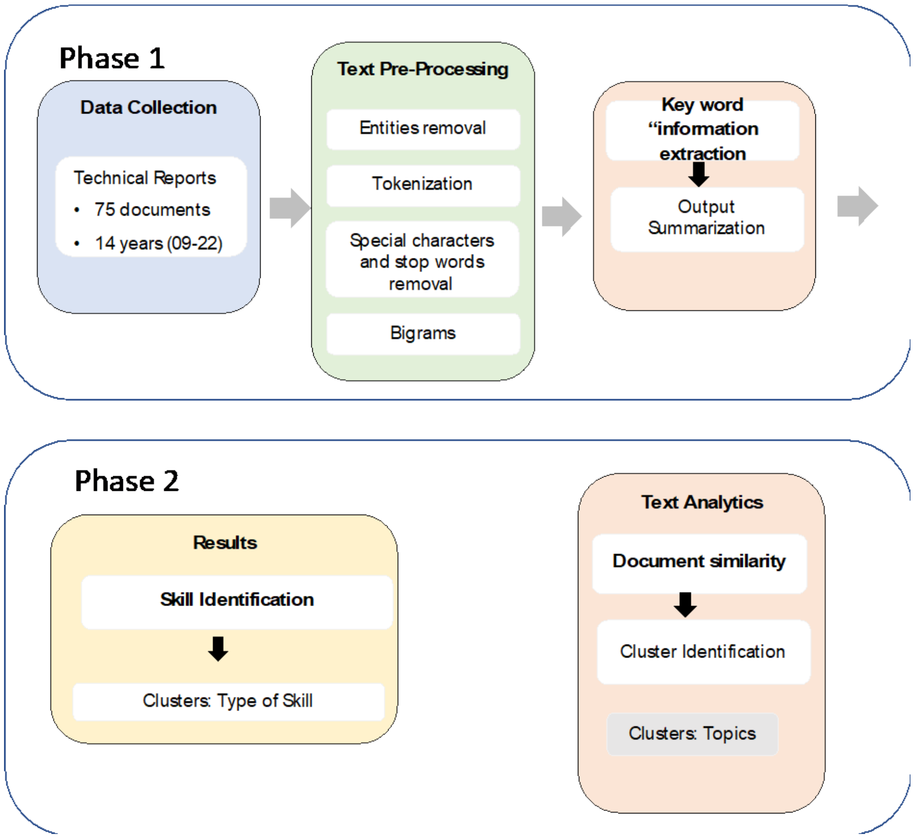


Figure 1. Skill Identification Framework Phases 1 and 2

The following sections describe the critical components of Phases 1 and 2, showcasing their functionality and applications.

Phase 1: Text Extraction and Summarization

In Phase 1, a robust toolset is provided for content extraction, keyword identification, summarization, and executive summary generation. This automation greatly enhances the efficiency and effectiveness of information management, enabling users to grasp crucial insights from extensive textual documents swiftly. The framework simplifies handling large volumes of text-based data, empowering users to make informed decisions, conduct comprehensive research, and produce concise executive summaries for effective communication and knowledge dissemination.

Code 1: Content Extraction and Identification of Key Information

Code 1 illustrates a Python-based solution (refer to Code 1 in Appendix A) for extracting content from PDF documents while identifying and highlighting specific keywords or phrases of interest. It employs NLP techniques to analyze the extracted text. The code functions as follows:

- Content Extraction: Utilizing the PyPDF2 library, the code extracts text from PDF documents, facilitating the processing of diverse textual content.
- Keyword Identification: Employing NLP, the code identifies keywords, lemmas, and stems of specified search terms within the extracted text and highlights them within the sentences, enabling users to identify relevant information swiftly.

- Output: Extracted sentences containing the specified keywords are presented with highlighted terms, and the code exports these sentences to a CSV file for further analysis or reference.

Code 2: Text Summarization for Executive Summary Generation

Code 2 presents a Python script (refer to Code 2 in Appendix A) for generating concise executive summaries from lengthy textual documents utilizing the "summarizer" library. The code operates as follows:

- Text Extraction: The code extracts text content from PDF files, preparing it for summarization.
- Text Summarization: Leveraging the "summarizer" library, the code generates an executive summary by selecting the most informative sentences from the document. Users can specify the desired summary length in sentences.
- Output: The executive summary is printed to the console, offering a succinct overview of the document's main points. This summary is then processed through AI-assisted technology to request an executive summary regarding skills and competencies.

▪ ***Phase 2.a***

Skill Identification and Analytics

Phase 2 of the framework, comprising Code 3 and Code 4, offers a versatile toolkit for document summarization and bigram extraction, essential for information organization, retrieval, and insight generation. These processes facilitate uncovering hidden relationships between documents, identifying shared bigrams indicative of common themes, and visualizing document similarity for effective content management. Codes 3 and 4 are pivotal components of this framework, demonstrating their functionality and applications in the context of skills, capabilities, and requirements.

Code 3: Document Similarity Analysis and Visualization

Code 3 (refer to Code 3 in Appendix A) provides a solution for comparing the similarity between PDF documents within a specified directory. The Python script performs the following tasks:

- Text Extraction: Extracts text content from multiple PDF documents in a designated directory.
- Cosine Similarity Calculation: Computes the cosine similarity between these documents using Term Frequency-Inverse Document Frequency (TF-IDF) vectors, quantifying the degree of textual resemblance.
- Heatmap Visualization: Generates a heatmap visually representing the similarity matrix to aid interpretation. The intensity of colors in the heatmap indicates the degree of similarity between pairs of documents, enabling users to identify clusters of related documents.
- Output: Presents results as a heatmap and a CSV file containing the similarity matrix for further analysis.

Code 4: Bigram Extraction and Document Clustering

Code 4 introduces a Python script for extracting bigrams (pairs of adjacent words) from PDF documents and their subsequent clustering (refer to Code 4 in Appendix A). The code performs the following tasks:

- Text Extraction: Extracts text content from multiple PDF documents in a specified directory.

- Text Preprocessing: Preprocesses the extracted text, including tokenization, removal of stopwords, and stemming using the Porter stemmer.
- Bigram Generation: Generates bigrams representing pairs of significant words from the preprocessed text, capturing meaningful word combinations for context and insights.
- TF-IDF Vectorization: Transforms bigrams into TF-IDF vectors, quantifying their importance in each document numerically.
- K-Means Clustering: Clusters documents with similar bigrams using the K-Means algorithm, grouping them into clusters.
- Output: Prints clusters of shared bigrams and associated documents. Additionally, AI-assisted technology structures unstructured text by providing insights into skills, talent, and capabilities. The AI-assisted technology is used to provide structure by requesting the following: Provide structure in terms of skill, talent, and, capabilities to the following unstructured text.

▪ ***Phase 2.b***

Network Analytics and Semantic Clustering

In this phase, the Louvain community detection algorithm is used as a pivotal tool in network analysis to identify communities or clusters within a given network based on the modularity of its structure (Puertas et al., 2021). This algorithm iteratively optimizes the network's modularity by dynamically reassigning nodes to different communities, forming cohesive and densely connected groups. In conjunction with Louvain community detection, NLP bigram extraction plays a crucial role in constructing the network. NLP techniques extract meaningful pairs of adjacent words, bigrams, from textual data. These bigrams serve as the nodes in the network, representing key concepts or entities derived from the text. The connections between nodes are established based on the co-occurrence of bigrams within a specified proximity, thereby capturing semantic relationships and associations in the text. By integrating Louvain community detection with NLP bigram extraction, a comprehensive semantic network can be developed, facilitating the exploration and analysis of complex textual data structures.

Framework Implementation and Discussion

Data Set

The dataset comprises a selection of 75 reports curated from an array of interactions with esteemed experts affiliated with the Systems Engineering Research Center (SERC) and the Acquisition Innovation Research Center (AIRC). These reports provide a rich spectrum of content, spanning sectors and perspectives, including but not limited to governmental insights, industry viewpoints encapsulated in position papers, and scholarly discourse in academic journal articles. A representative sample is included in Appendix A for a glimpse into these reports. If needed, the interested reader may request access to the complete archive.

Skill Extraction Results

The skill extraction process yielded results from all 75 reports, each of which underwent integration into the semi-automated framework, resulting in the generation of executive reports. Figure 2 illustrates the implementation of Phase 1 of the framework using a specific report as an example, while Table 1 presents the corresponding summary output. Notably, the framework successfully generated 60 executive summaries, although 20 initial documents provided minimal or no information regarding skill sets. These executive summaries can be provided upon request.

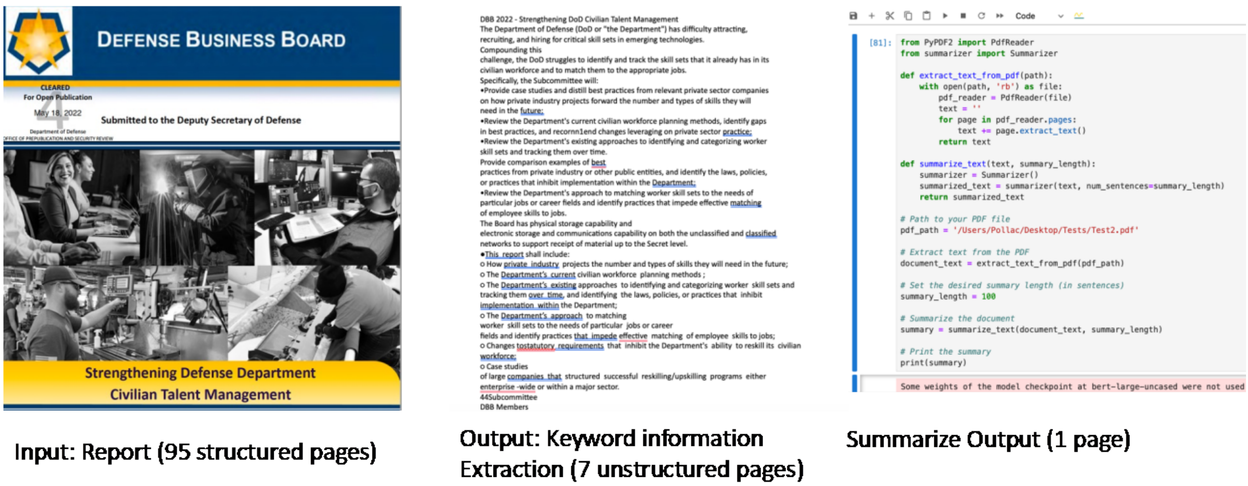


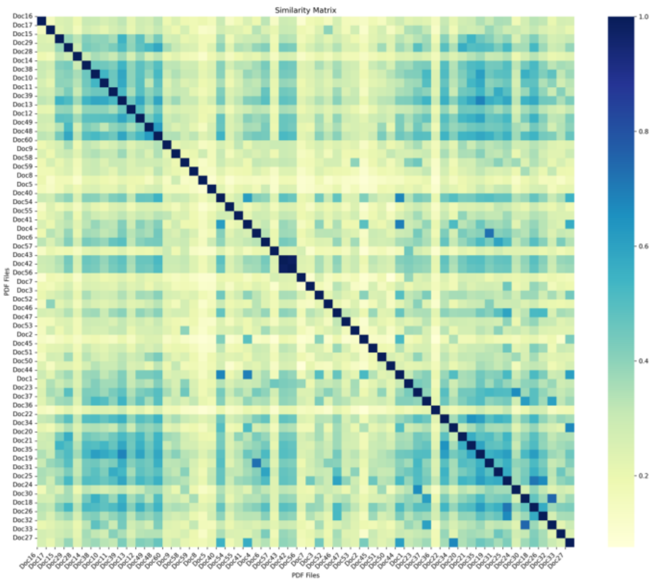
Figure 2. Defense Business Board Document Phase 1 Example

Table 1. Executive Summary for Defense Business Board Document

1. Talent Acquisition and Recruitment: Skills in attracting and recruiting professionals with critical skill sets in emerging technologies. This includes expertise in sourcing candidates, conducting interviews, assessing qualifications, and employing effective recruitment strategies.	6. Change Management: Proficiency in managing change within the organization to facilitate the adoption of new talent management practices. This includes communication, stakeholder engagement, and developing strategies to address resistance or challenges related to implementing new approaches to talent management.
2. Workforce Planning: Skills in strategic workforce planning to anticipate and align human capital needs with organizational goals. This involves analyzing current and future skill requirements, identifying gaps, and developing plans to address those gaps through recruitment, training, or other talent management initiatives.	7. Data Management and Analysis: Competence in data management and analysis to support talent management decisions. This includes using technology platforms and data lakes to track and analyze job-related data, employee skills, and workforce trends, enabling informed decision-making and proactive planning.
3. Skill Set Identification and Tracking: Competence in identifying and categorizing worker skill sets, as well as establishing systems to track and update these skills over time. This includes using technology and data analysis to monitor skill inventories, assess skill gaps, and ensure accurate matching of employee skills to job requirements.	8. Collaboration and Relationship Building: Skills in building partnerships and collaborative relationships with stakeholders both within and outside the organization. This includes fostering cooperation between different departments, leveraging external expertise, and engaging with private industry partners to exchange knowledge and best practices.
4. Comparative Analysis and Benchmarking: Skills in conducting comparative analysis between the DoD's talent management practices and those of private sector companies or other public entities. This entails identifying best practices, gaps, and areas for improvement, and making recommendations for adapting private industry practices to enhance talent management in the DoD.	9. Talent Development and Upskilling: Abilities in designing and implementing talent development programs that enhance employees' skills and promote lifelong learning. This involves creating learning opportunities, providing access to training resources, and encouraging employees to expand their knowledge and expertise in line with organizational needs.
5. Skill Matching and Job Alignment: Abilities in matching worker skill sets to the needs of specific jobs or career fields within the DoD. This requires understanding the knowledge, experience, and competencies required for different roles and effectively aligning employees' skills to maximize their contributions and job satisfaction.	10. Knowledge of Emerging Technologies: Understanding and awareness of emerging technologies relevant to the defense sector. This includes staying abreast of advancements in areas such as cybersecurity, artificial intelligence, data analytics, robotics, and other emerging fields that impact DoD operations and require specialized skill sets.

Skill Identification

After acquiring the executive summaries, Phase 2.a involved creating a similarity matrix for all documents, as demonstrated in Figure 3, and extracting pertinent bigrams from document clusters. The task of identifying skill sets was completed utilizing AI-assisted technology, with ChatGPT being the designated tool for this task, as outlined in Table 2. The final output encompasses the comprehensive compilation of skill sets from all 15 clusters (available upon request, given its size). It is essential to emphasize that labeling these skill sets is AI-generated and may require further refinement based on a thorough understanding of DoD requirements.



- Cluster 9 - Shared Bigrams: aspect life, skill item, skill framework, skill pattern, execut summari, scale internet, conclus studi, futur research, item clariti, method employ, comprehens set, test internet, develop reliabl, studi seek, interview conduct, et al, set reliabl, internet skill, differ popul, valuabl futur, type engag, find three, occup group, compar skill, base find, categori contain, studi aim, better understand, level measur, framework base, valid reliabl, internet use, systemat literatur, reliabl valid, polici evalu, report emphas, goe beyond, skill goe, aim develop, pattern relat, van deursen, measur internet, emphas import, five main, coher studi, import internet, reliabl measur, relat differ, inform navig, propos scale, literatur review, educ level, studi found, skill relat, categori oper, report focus, aspect internet, life outcom, impact variou, research better, use valid, holist view, across differ, measur skill, summari provid, examin extern, seek identifi, measur allow, reliabl cognit, identifi comprehens, digit literaci, age educ, inclus everyday, creativ mobil, skill scale, skill execut, skill level, main skill, differ aspect, three basic, methodolog studi, evalu propos, clariti coher, skill understand, develop internet, difficulti use, skill digit, consist skill, applic survey, literaci digit, propos skill, skill categori, technic aspect, user engag, investig skill, summari van, impact user, inclus effort, digit inclus, valid measur, life background, navig social, background focu, popul group, framework includ, level occup, view skill, develop set, outcom propos, valid report, found consist, relat gender, group conclus, valid internet, outcom studi, al measur, found item, employ investig, skill measur, understand relat, identifi studi, project report, basic method, focus measur, engag outcom, oper inform, inclus impact, specif skill, research intervent, propos framework, ensur accuraci, everyday life, contain specif, extern valid, conduct evalu, conduct systemat, valid compar, survey perform, accuraci understand, perform test, studi develop, variou aspect, evalu relat, item need, focu report, differ type, need revis, intervent polici, skill methodolog, use applic, social creativ, inclus propos, allow research, report examin, relat digit, engag impact, includ five, skill difficulti, deursen et, mobil skill, revis ensur, skill valuabl, understand digit, aspect studi, cognit interview, gender age, beyond technic, propos holist, understand extern, scale across, skill impact, review identifi, studi conduct, group found, framework propos, provid overview, overview project, impact digit Related Documents: - Doc59.pdf

Figure 3. Similarity Matrix for All Executive Summaries of Phase 1 and Cluster 9

Network Analytics

After completing Phase 1 (see Figure 4) of the framework and to construct a semantic network suitable community detection, the following steps involve defining the nodes of the network, which are determined as the most frequent bigrams within the corpus. Following node selection, the network's links are established based on the co-occurrence of bigrams within a proximity window of size 10. These links are then weighted according to the frequency of co-occurrences. Once the semantic network is established, Louvain community detection is applied to identify semantic clusters within the network. The Louvain algorithm operates iteratively, beginning with small communities and gradually adjusting the modularity by adding or removing nodes from communities. This process continues until an optimal modularity score is achieved. Nodes with higher modularity scores, indicative of denser clusters, are grouped together, thus forming distinct semantic clusters within the network.

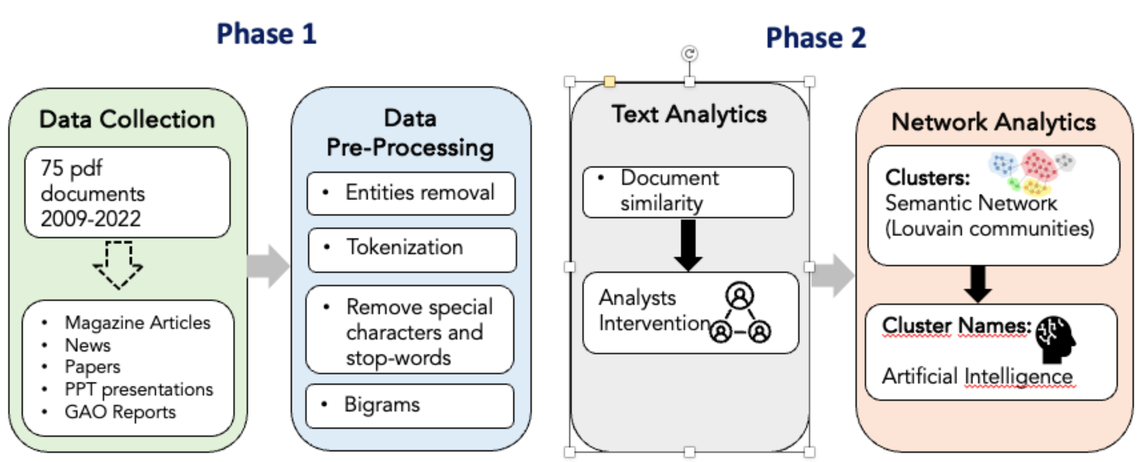


Figure 4. Network Analytics Stage within Framework Phases 1 and 2

The network analysis of the implemented Louvain algorithm on the series of documents generated a network consisting of 318 nodes and 4,743 edges, indicating a density of 0.09410. The network exhibited a transitivity of 0.37795 and an average clustering coefficient of 0.5030, suggesting a moderate level of clustering within the network. Furthermore, nodes with higher centrality, such as "United_States," "Department_defense," and "National_security," emerged as prominent entities, highlighting their significance in the network structure. Additionally, nodes with higher closeness, including "Artificial_Intelligence," "Big_data," and "Data_science," were identified, underscoring their pivotal role in facilitating efficient information flow within the network. These findings provide valuable insights into the key entities and their interconnectedness within the document network, shedding light on the underlying themes and relationships present in the dataset. Figure 5 illustrates the final network obtained along with the communities identified.

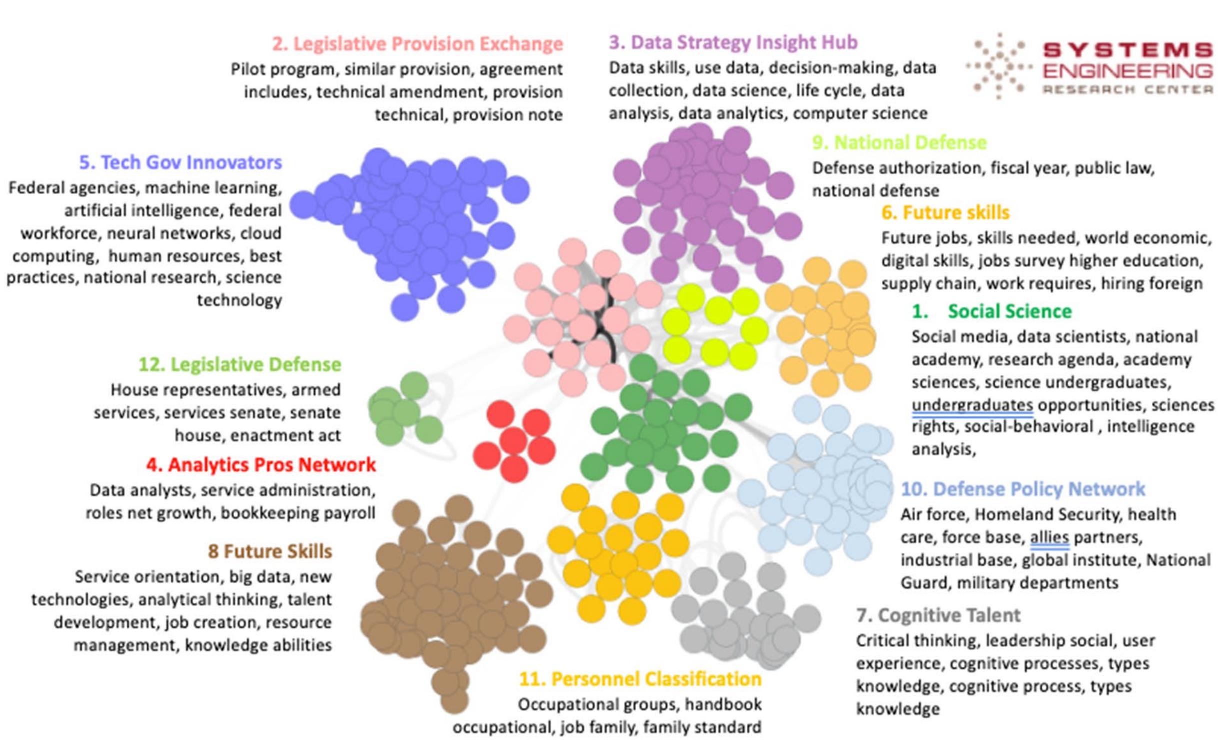


Figure 5. Skill Communities Identified

Conclusions and Future Work

The proposed framework has adeptly addressed the core research inquiries guiding this study. Initially, it confronts the intricate task of skill identification within the DoD workforce by harnessing the capabilities of NLP algorithms. Through meticulous analysis of a diverse corpus comprising over 80 documents from industry, government, and academic reports, the framework seamlessly automates the detection of pivotal skills. This approach furnishes valuable insights into the indispensable skill sets for tackling global challenges and proficiently managing the DoD workforce, bolstering the organization's preparedness and robustness.

Moreover, the framework propels the realm of decision support within the DoD by capitalizing on NLP analysis to empower decision-makers. By distilling actionable insights and recommendations from the extensive document analysis, it equips DoD leadership with the requisite knowledge to make judicious decisions concerning talent acquisition, training initiatives, and the mitigation of skill disparities. In an epoch characterized by dynamic global challenges, the framework's contributions transcend mere automation; it emerges as a strategic apparatus for augmenting the agility and efficacy of the DoD, thereby enhancing the organization's ability to navigate evolving complexities with acumen and foresight.

Promising opportunities for further research and development are evident in several domains, with Skill Mapping emerging as a crucial area for advancement within the DoD. Future endeavors in this field aim to develop an intelligent system capable of mapping identified skills to specific job roles and career pathways. This innovative approach has the potential to revolutionize talent allocation, enabling the DoD to match personnel skills with roles and development trajectories precisely, thereby enhancing organizational agility and preparedness.

Moreover, the scope of NLP-driven talent management extends to Predictive Analysis, offering fertile ground for exploration. Future initiatives can focus on forecasting forthcoming skill requirements by harnessing the power of NLP-driven predictive models. These models would facilitate proactive workforce planning and readiness by integrating insights from emerging trends, technological advancements, and evolving operational needs. Such predictive capabilities are poised to significantly impact the DoD's ability to anticipate skill demands, stay ahead of evolving challenges, and cultivate a workforce equipped to navigate the dynamic demands of a rapidly changing global landscape. In summary, the realm of NLP-based talent management holds immense promise, and these avenues for further research and development serve as guiding lights toward fostering a more agile and effective DoD.

References

Allioui, H., & Mourdi, Y. (2023). Unleashing the potential of AI: Investigating cutting-edge technologies that are transforming businesses. *International Journal of Computer Engineering and Data Science (IJCEDS)*, *3*(2), 1–12. https://ijceds.com/ijceds/article/view/59

Caratozzolo, P., Alvarez-Delgado, A., & Rodriguez-Ruiz, J. (2023). Applications of natural language processing for industry 4.0 skills development. *2023 Future of Educational Innovation-Workshop Series Data in Action*, 1–9. https://doi.org/10.1109/IEEECONF56852.2023.10104796

Gomathi, S., Rajeswari, A., & Kadry, S. (2023). Emerging HR practices—digital upskilling: A strategic way of talent management and engagement. In *Disruptive artificial intelligence and sustainable human resource management* (1st ed.). River Publishers. https://www.taylorfrancis.com/books/edit/10.1201/9781032622743/disruptive-artificial-intelligence-sustainable-human-resource-management-anamika-pandey-balamurugan-balusamy-naveen-chilamkurti?refId=85c9b79f-adb5-461d-a002-77c14fe3d195&context=ubx

Mohanty, S., Behera, A., Mishra, S., Alkhayyat, A., Gupta, D., & Sharma, V. (2023). Resumate: A prototype to enhance recruitment process with NLP based resume parsing. *2023 4th International Conference on Intelligent Engineering and Management (ICIEM)*, 1–6. https://doi.org/10.1109/ICIEM59379.2023.10166169

Puertas, E., Moreno-Sandoval, L. G., Redondo, J., Alvarado-Valencia, J., & Pomares-Quimbaya, A. (2021). Detection of sociolinguistic features in digital social networks for the detection of communities. *Cognitive Computation*, *13*, 518–537. https://doi.org/10.1007/s12559-021-09818-9

Vanetik, N., & Kogan, G. (2023). Job vacancy ranking with sentence embeddings, keywords, and named entities. *Information*, *14*(8), 468. https://doi.org/10.3390/info14080468

Appendix A. Selected Reports

Presidents Management Agenda, Federal Data Strategy, Curated Data Skills Catalog, November 2020. Accessed: https://strategy.data.gov/action-plan/

World Economic Forum, The future of Jobs Report, October 2020,

Accessed: https://www.weforum.org/publications/the-future-of-jobs-report-2020/

Jacobson, S. Maximizing the data Literacy of the Airforce Contracting Workforce, MBA Professional Project, Naval Postgraduate School, December 2021.

The White House, National Security Strategy, October 2022.

Accessed: https://www.whitehouse.gov/wp-content/uploads/2022/10/Biden-Harris-Administrations-National-Security-Strategy-10.2022.pdf

U.S. Department of Defense, 2022 National Defense Strategy, October 2022.

Accessed: https://www.defense.gov/News/News-Stories/Article/Article/3202438/dod-releases-national-defense-strategy-missile-defense-nuclear-posture-reviews/#:~:text=The%202022%20National%20Defense%20Strategy,and%20partners%20on%20shared%20objectives.

U.S. Department of Defense, National Defense Science and Technology Strategy 2023.

Accessed: https://media.defense.gov/2023/May/09/2003218877/-1/-1/0/NDSTS-FINAL-WEB-VERSION.PDF

Adams NE. Bloom's taxonomy of cognitive learning objectives. J Med Libr Assoc. July 2015;103(3):152-3. doi: 10.3163/1536-5050.103.3.010. PMID: 26213509; PMCID: PMC4511057.

Anderson, L. W. and Krathwohl, D. R., et al (Eds..) A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Allyn & Bacon. Boston, MA (Pearson Education Group) 2021

Scherger Group, Future Workforce 2025, September 2019.

Accessed: https://theforge.defence.gov.au/article/future-workforce-2025-scherger-group

McKinsey Global Institute. Skill Shift Automation and the Future of the Workforce. May 2018.

Accessed: https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce

Future Skills Council. Canada – A Learning Nation: A Skilled, Agile Workforce Ready to Shape the Future. November 2020. ISBN: 978-0-660-35859-8

Deloitte Insights. The skills-based organization: A new operating model for work and the workforce. September 2022.

Accessed: https://www2.deloitte.com/us/en/insights/topics/talent/organizational-skill-based-hiring.html

Gehlhaus, D., Ryseff, J. and Corrigan, J. The Race for U.S. Technical Talent: Can the DOD and DIB Compete? Center for Security and Emerging Technology. August 2023. DOI: 10.51593/20210074

Defense Business Board. Strengthening Defense Department Civilian Talent Management. May 2022

Accessed: https://dbb.defense.gov/Portals/35/Documents/Reports/2022/DBB%20FY22-03%20Talent%20Management%20Study%20Report%2018%20Aug%202022%20-%20CLEARED.pdf

Defense Civilian Personnel Advisory Service. Strategic Workforce Planning Guide. May 2019

Accessed: https://www.dcpas.osd.mil/sites/default/files/DoD%20Strategic%20Workforce%20Planning%20Guide%20-%2030May2019.pdf

Appendix B. Code

Code 1

```
import os
import ssl
import certifi
import PyPDF2
import nltk
from nltk.tokenize import sent_tokenize, word_tokenize
from nltk.stem import WordNetLemmatizer, PorterStemmer
from termcolor import colored
import csv
# Set the SSL certificate verification
ssl._create_default_https_context = ssl.create_default_context(cafile=certifi.where())
def load_pdf(file_path):
    try:
        with open(file_path, "rb") as file:
            reader = PyPDF2.PdfReader(file)
            text = ""
            for page in reader.pages:
                text += page.extract_text()
            return text
    except FileNotFoundError:
        print("File not found.")
        return None
    except PyPDF2.PdfReadError:
        print("Invalid PDF file.")
        return None
def extract_sentences_with_words(text, words):
    sentences = sent_tokenize(text)
```

```
    sentences_with_words = []
    lemmatizer = WordNetLemmatizer()
    stemmer = PorterStemmer()
    # Extract sentences with the specified words
    for sentence in sentences:
        tokens = word_tokenize(sentence)
        lemmas = [lemmatizer.lemmatize(token) for token in tokens]
        stems = [stemmer.stem(token) for token in tokens]

        # Prepare the set of unique word forms (word, lemma, and stem) to search for
        search_words = set()
        for word in words:
            search_words.add(word)
            search_words.add(lemmatizer.lemmatize(word))
            search_words.add(stemmer.stem(word))
        if any(token in search_words for token in tokens) or any(lemma in search_words for lemma
in lemmas) or any(stem in search_words for stem in stems):
            highlighted_sentence = highlight_words(sentence, tokens, lemmas, stems,
search_words)
            sentences_with_words.append(highlighted_sentence)
    return sentences_with_words
def highlight_words(sentence, tokens, lemmas, stems, search_words):
    highlighted_sentence = sentence
    for token in tokens:
        if token in search_words:
            highlighted_sentence = highlighted_sentence.replace(token, colored(token, 'yellow'))
    for lemma in lemmas:
        if lemma in search_words:
            highlighted_sentence = highlighted_sentence.replace(lemma, colored(lemma, 'yellow'))
    for stem in stems:
        if stem in search_words:
            highlighted_sentence = highlighted_sentence.replace(stem, colored(stem, 'green'))
    return highlighted_sentence
# Set the PDF file path
pdf_file = '/Users/ Tests/Test1.pdf'
```

```
# Load the PDF file
text = load_pdf(pdf_file)
if text:
    # Specify the list of words you want to search for
    words = ["skill", "competencies", "capability", "talent"]
    # Extract sentences with the specified words
    sentences = extract_sentences_with_words(text, words)
    # Print the extracted sentences with highlighted words, lemmas, and stems
    if sentences:
        print("Sentences with the specified words:")
        for sentence in sentences:
            print(sentence)
    else:
        print("No sentences found with the specified words.")
    # Save the output to a CSV file
    output_file = 'output.csv'
    with open(output_file, 'w', newline='', encoding='utf-8') as csvfile:
        writer = csv.writer(csvfile)
        writer.writerow(['Sentences with the specified words:'])
        writer.writerows([[sentence] for sentence in sentences])
        writer.writerow(['No sentences found with the specified words.'])
```

Code 2

```
from PyPDF2 import PdfReader
from summarizer import Summarizer
def extract_text_from_pdf(path):
    with open(path, 'rb') as file:
        pdf_reader = PdfReader(file)
        text = ''
        for page in pdf_reader.pages:
            text += page.extract_text()
        return text
def summarize_text(text, summary_length):
    summarizer = Summarizer()
    summarized_text = summarizer(text, num_sentences=summary_length)
    return summarized_text
```

```
# Path to your PDF file
pdf_path = '/Users/ Tests/Test2.pdf'
# Extract text from the PDF
document_text = extract_text_from_pdf(pdf_path)
# Set the desired summary length (in sentences)
summary_length = 100
# Summarize the document
summary = summarize_text(document_text, summary_length)
# Print the summary
print(summary)
```

Code 3

```
import os
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from PyPDF2 import PdfReader
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
def read_pdf(file_path):
    with open(file_path, 'rb') as file:
        pdf_reader = PdfReader(file)
        text = ""
        for page in pdf_reader.pages:
            text += page.extract_text()
        return text
def compare_similarity(documents_dir):
    # Get a list of PDF files in the given directory
    pdf_files = [file for file in os.listdir(documents_dir) if file.endswith('.pdf')]
    # Read and preprocess the content of each PDF document
    texts = []
    for pdf_file in pdf_files:
        file_path = os.path.join(documents_dir, pdf_file)
        text = read_pdf(file_path)
        texts.append(text)
    # Compute TF-IDF vectors for the documents
```

```
    vectorizer = TfidfVectorizer()
    tfidf_matrix = vectorizer.fit_transform(texts)
    # Calculate the cosine similarity matrix
    similarity_matrix = cosine_similarity(tfidf_matrix)
    # Convert similarity matrix to DataFrame for better representation
    similarity_df = pd.DataFrame(similarity_matrix, columns=pdf_files, index=pdf_files)
    # Extract file names without extensions
    pdf_names = [os.path.splitext(pdf_file)[0] for pdf_file in pdf_files]
    # Set up the figure and axis
    plt.figure(figsize=(16, 13))
    ax = sns.heatmap(similarity_df, annot=False, cmap='YlGnBu', fmt=".2f")
    # Customize the plot
    ax.set_title("Similarity Matrix")
    ax.set_xlabel("PDF Files")
    ax.set_ylabel("PDF Files")
    plt.xticks(range(len(pdf_names)), pdf_names, rotation=45, ha='right')
    plt.yticks(range(len(pdf_names)), pdf_names, rotation=0)
    # Save the plot as an image (JPEG or PNG)
    output_image_path = os.path.join(documents_dir, "similarity_matrix.png")
    plt.tight_layout()
    plt.savefig(output_image_path, dpi=300)
    plt.show()
  # Output similarity matrix to a CSV file
    similarity_df.to_csv(os.path.join(documents_dir, "similarity_matrix.csv"))
if __name__ == "__main__":
    # Replace 'path/to/directory' with the path to your directory containing the PDF files
    documents_directory = '/Users /Tests'
    compare_similarity(documents_directory)
```

Code 4:

```
import os
import ssl
import nltk
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from sklearn.feature_extraction.text import TfidfVectorizer
```

```
from sklearn.cluster import KMeans
import numpy as np
from PyPDF2 import PdfReader

# Configure SSL context to bypass certificate verification for NLTK downloads
try:
    _create_unverified_https_context = ssl._create_unverified_context
except AttributeError:
    pass
else:
    ssl._create_default_https_context = _create_unverified_https_context

# Download necessary NLTK resources
nltk.download('punkt')
nltk.download('stopwords')

# Function to read text from PDF
def read_pdf(file_path):
    with open(file_path, 'rb') as file:
        pdf_reader = PdfReader(file)
        text = ""
        for page in pdf_reader.pages:
            text += page.extract_text()
        return text

# Function to preprocess the text and generate bigrams
def preprocess_text(text):
    # Tokenize the text into words
    words = nltk.word_tokenize(text.lower())

    # Remove stopwords and non-alphabetic characters from words
    stop_words = set(stopwords.words('english'))
    words = [word for word in words if word.isalpha() and word not in stop_words]

    # Stemming using Porter stemmer
```

```
    stemmer = PorterStemmer()
    words = [stemmer.stem(word) for word in words]

    # Generate bigrams
    bigrams = list(nltk.bigrams(words))
    bigrams = [" ".join(bigram) for bigram in bigrams]

    return bigrams

# Specify the directory containing PDF documents
documents_directory = '/Users/Tests'

# Get a list of PDF files in the given directory
pdf_files = [os.path.join(documents_directory, file) for file in os.listdir(documents_directory) if
file.endswith('.pdf')]

# Read and preprocess the content of each PDF document
documents = [read_pdf(file_path) for file_path in pdf_files]

# Preprocess the documents and generate bigrams
preprocessed_documents = [preprocess_text(doc) for doc in documents]

# Calculate TF-IDF vectors for bigrams
vectorizer = TfidfVectorizer()
tfidf_vectors = vectorizer.fit_transform([" ".join(bigrams) for bigrams in
preprocessed_documents])

# Clustering using K-Means
num_clusters = 15  # Change this value based on the number of clusters you want
kmeans = KMeans(n_clusters=num_clusters, random_state=42)
clusters = kmeans.fit_predict(tfidf_vectors)

# Create clusters of shared bigrams
cluster_bigrams = {}
for i, cluster_label in enumerate(clusters):
```

```
    if cluster_label not in cluster_bigrams:
        cluster_bigrams[cluster_label] = set()
    bigrams = preprocessed_documents[i]
    cluster_bigrams[cluster_label].update(bigrams)
# Create a dictionary to store documents corresponding to each cluster
cluster_documents = {}
for i, cluster_label in enumerate(clusters):
    if cluster_label not in cluster_documents:
        cluster_documents[cluster_label] = []
    cluster_documents[cluster_label].append(pdf_files[i])

# Print the clusters and shared bigrams along with associated documents
for cluster_label, bigrams in cluster_bigrams.items():
    print(f"Cluster {cluster_label + 1} - Shared Bigrams:")
    print(", ".join(bigrams))
    print("Related Documents:")
    for document_path in cluster_documents[cluster_label]:
        print(f"- {os.path.basename(document_path)}")
    print("---")
```

Planning for AI Sustainment: A Methodology for Maintenance and Cost Management

MAJ Iain Cruickshank—is a Functional Area 49 (Operations Research/Systems Analysis) officer in the U.S. Army. He is currently a senior research scientist at the Army Cyber Institute. He has previous assignments with the Army's Artificial Intelligence Integration Center, the 780th Military Intelligence Brigade, and the 101st Airborne Division. He holds a PhD in Societal Computing from Carnegie Mellon University, which was obtained as a National Science Foundation Graduate Research Fellow, and an MS in Operations Research from the University of Edinburgh, which was obtained as a Rotary Ambassadorial Scholar. [iain.cruickshank@westpoint.edu]

MAJ Shane Kohtz—is a Functional Area 51 (Acquisition) officer in the U.S. Army. He is currently a cyber research manager at the Army Cyber Institute. He has previous assignments with the Missile Defense Agency, the Army's Program Executive Office Intelligence Electronic Warfare & Sensors, the 1st Infantry Division, and the 101st Airborne Division. He holds an MBA from the Naval Postgraduate School with a focus in Systems Acquisition Management. He is a member of the Army Acquisition Corps and holds a DAWIA Advanced certification in program management. [shane.kohtz@westpoint.edu]

Abstract

The sustainment requirements of Artificial Intelligence (AI)–enabled systems are largely unexplored within the Department of Defense's programs of record (POR). Many programs often overlook maintenance needs for AI systems, extending beyond base hardware or software upkeep. However, prior research indicates a distinctive maintenance requirement for the machine learning models that power AI-enabled systems, and outlines strategies for planning and integrating AI maintenance into product support (Cruickshank & Kohtz, 2023). Notably, the adoption of industry best practices, program maintenance considerations, and machine learning operations (MLOps) are crucial for crafting an AI system's sustainment strategy. This research builds upon the existing framework to further comprehend the extent of preventative and routine maintenance required by an AI-enabled system. We specifically investigate the degree of maintenance or "touch-time" needed to sustain a system's machine learning model(s). By examining a typical year of operations and sustainment for an AI-enabled computer vision system, we highlight primary maintenance considerations (i.e., maintenance tasks, task difficulty, and task frequency) and propose a method to estimate these factors. We then apply varying levels of maintenance based on organic, hybrid, or contractor logistics support to fully comprehend the sustainment costs. Our research offers a robust framework for program offices to more accurately predict initial and ongoing operation and sustainment costs when conducting a business case analysis. This will enable the selection of the most cost-effective sustainment strategy for a POR that intends to use any AI-enabled system.

Introduction

As technology advances on the battlefield, the task of achieving superiority over adversaries becomes increasingly challenging. The introduction of Artificial Intelligence (AI)–enabled systems with machine learning (ML) models holds promise in offering a competitive edge. However, this technology is still in its early stages of development and deployment across the armed forces. Understanding the full scope and sustainment requirements of integrating this technology into both existing and new weapon systems presents a significant challenge. This research aims to bridge the gap in understanding and planning for future product support strategies. An AI-enabled system demands additional maintenance beyond the typical hardware and software upkeep observed in existing systems. It is imperative to recognize the necessity of treating ML models as systems within systems, marking a paradigm shift essential for determining appropriate sustainment strategies. The background of this paper delves into previous research inputs that inform a product support strategy for an AI-enabled system.

Our research has analyzed the tasks necessary for ML maintenance and identified those tasks that service members are capable of performing and identified those maintenance tasks that are best executed by contractors due to their advanced technical requirements. As a result, we recommend a hybrid approach to a product support strategy, combining contractor and service support, for any AI-enabled system. This research implements the aforementioned approach to develop labor hour requirements for specific maintenance costs and a cost estimating model for two different maintenance strategy options: contractor-only and a hybrid of contractor and service support. These approaches are essential for enabling decision-makers to discern the most cost-effective approach in the early planning stages of a product support strategy over a typical 20-year service life cycle for a program of record (POR).

Background

This section provides crucial background information pertinent to the proposed maintenance cost models presented in this paper. Specifically, it covers the maintenance requirements of AI-enabled systems, the MLOps paradigm concerning the utilization of ML models in real-world systems, and previous research regarding the maintenance of ML models in military AI-enabled systems.

AI-enabled systems and their maintenance

AI-enabled systems, akin to any technological apparatus, require maintenance. Such systems comprise traditional software and, potentially, hardware, contingent upon the system's purpose, in addition to AI components. The AI components often rely on various hardware and software dependencies, commonly referred to as a "stack" (Moore et al., 2018). Figure 1 illustrates the AI stack. The critical elements of AI components, which render the entire system AI-enabled, are the ML models. These models empower the system to execute automated behaviors and tasks that typically demand human-level perception or reasoning, serving as the "brain" of the AI-enabled system. Just like every other component of an AI-enabled system, these ML models also require maintenance.

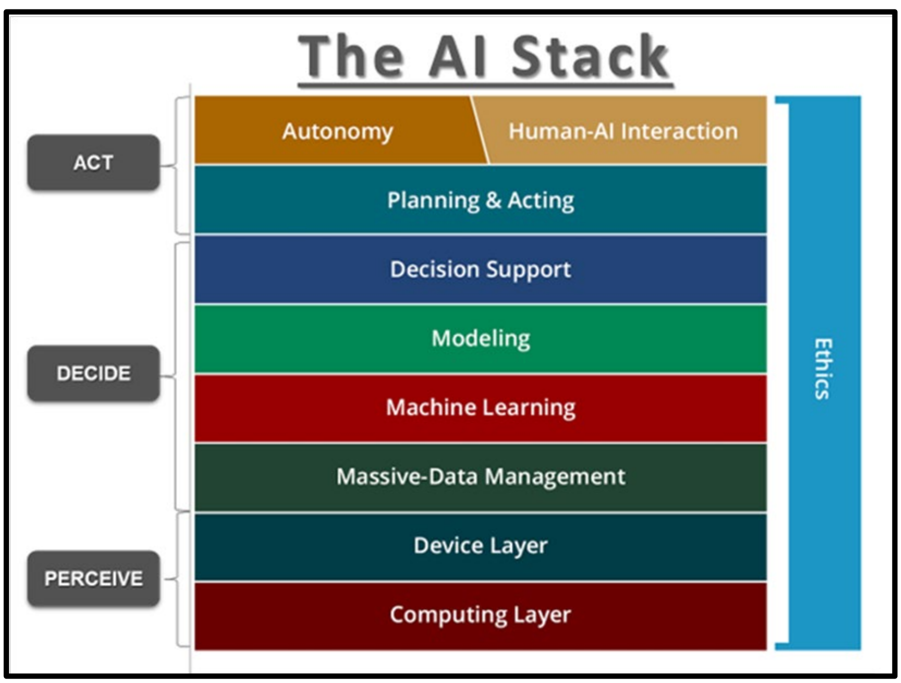


Figure 1. Carnegie Mellon University's AI Stack, Depicting the Necessary Components of an AI-Enabled System (Moore, 2018)

Despite their potential, ML models still encounter several issues that necessitate frequent maintenance. ML models inherently learn correlations useful for specific tasks from the data presented to them. Consequently, performance issues may arise if the data during utilization differs from the training data (i.e., out-of-domain data problem; Patruno, 2019). For instance, a computer vision ML model designed to detect specific vehicles from a ground perspective may fail when confronted with differences in background or biome between its training data and operational environment (e.g., urban versus rural settings). Additionally, ML models can suffer from issues like model drift (Talby, 2018), data drift (Evidently AI, 2021), concept drift (Patruno, 2019), or changes in hardware such as sensors. All of these changes, which generally would not perturb a change in a human's task performance, significantly impact ML model performance. Furthermore, ML models can be directly targeted through adversarial ML, resulting in substantial degradation of model performance (Talby, 2018). Notably, many of these issues are unique to ML and ML-enabled systems; changes such as alterations to image backgrounds do not affect the hardware or software of traditional digital systems. Hence, ML models entail their own inherent issues necessitating maintenance beyond that required for traditional hardware and software systems.

While ML models face several issues that can greatly affect their performance, addressing these issues often demands fewer resources and expertise compared to the initial development of the ML model. Maintaining ML models deployed in real-world settings—referred to as model deployment—can typically be managed through a suite of updating and monitoring processes, collectively forming part of the industrial ML paradigm known as MLOps (Treveil et al., 2020). At its core, MLOps encompasses a set of practices aimed at operationalizing ML systems (Treveil et al., 2020). Figure 2 illustrates the core components and relationships of MLOps. Although the principles and practices of MLOps remain an active area of research, three practices integral to MLOps include data and model monitoring in production, continuous model updates in response to changes, and integrating model maintenance into model operation (Treveil et al., 2020). These practices are essential for organizations and businesses to utilize ML models despite their inherent issues. Thus, effective employment of ML models in real-world and production systems within the MLOps paradigm requires the implementation of appropriate tools and practices for monitoring ML models and their data, as well as procedures for updating ML models as close to operational use as possible.

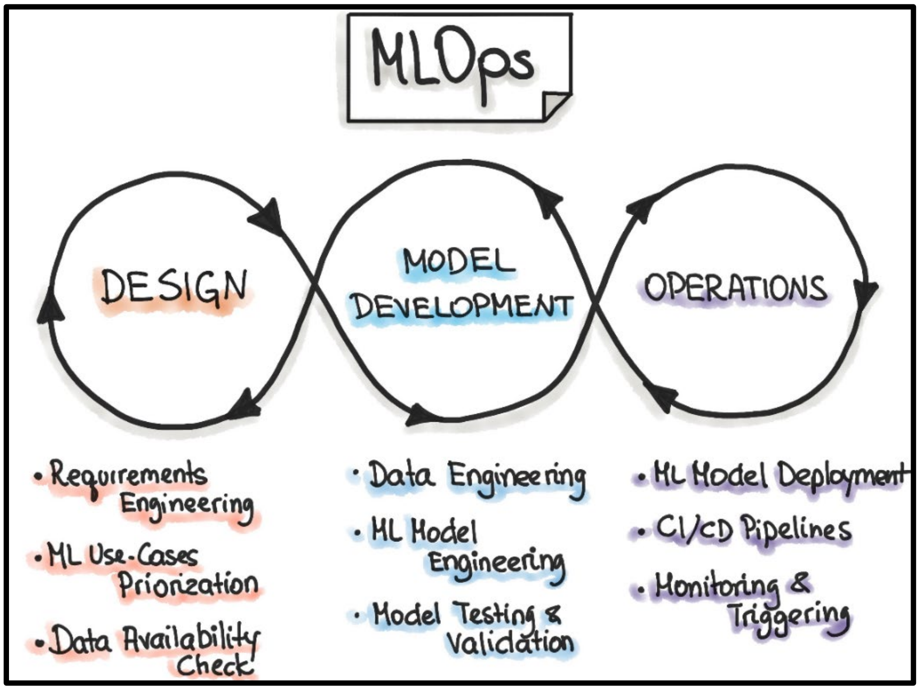


Figure 2. Core Components of MLOps and Their Relationships (Visengeriyeva et al., 2023)

Of particular significance within the MLOps paradigm is *model retraining*. Ideally, model retraining involves rerunning all steps required to train an ML model with a new dataset, necessitating changes only to the model's weights, not its code (Patruno, 2019). This form of maintenance typically occurs whenever the data changes and an updated training dataset becomes available (Evidently AI, 2021). Hence, this maintenance generally takes two forms: periodic and dynamic (Evidently AI, 2021). Periodic retraining involves anticipated changes in data, such as quarterly or yearly shifts in business practices, while dynamic retraining occurs whenever there are changes in the data generation process, such as collecting data in an adversarial environment (e.g., detecting credit fraud) or in a naturally dynamic process (e.g., labeling objects in imagery). The frequency of dynamic retraining can vary considerably depending on the ML application; some ML models require daily updates, while others may only need monthly or yearly updates (Evidently AI, 2021). Furthermore, depending on the system in which the ML model is used, retraining may also involve updating models across several devices, wherein the new model is pushed or flashed onto those devices; once retrained, the model must be reintegrated back into the AI-enabled system. Regardless of the frequency of ML model retraining, all experts agree that this process is essential for any ML-enabled system. Thus, model retraining is a necessary component of any ML model and may need to occur as frequently as daily.

Maintenance Considerations for Military AI-Enabled Systems

In contemplating maintenance strategies for AI-enabled systems, several key considerations emerge. These considerations are pivotal in guiding program offices during the product support business case analysis (PS BCA), which informs both the product support strategy (PSS) and life cycle sustainment plan (LCSP; DoD, 2022b). The PS BCA assesses alternative sustainment options, including organic, contractor, or a hybrid mix of support, thereby informing the program's sustainment strategy (DoD, 2014). Notably, approximately 85% of sustainment costs are established during the requirement setting phase (Schinasi, 2003). Therefore, understanding the requirements and maintenance demands of AI/ML systems is crucial during the strategy development phase to facilitate effective planning and budgeting for sustainment. This encompasses the maintenance of ML models in addition to the hardware and software components underlying the AI stack, which are essential for ML model operation within the system.

Various paradigms exist for approaching maintenance of ML models, akin to sustaining other components of a system, offering both contract and organic service support alternatives. At one end of the spectrum lies exclusive contract-based maintenance for ML models. Under this arrangement, contractors assume full responsibility for all aspects of model maintenance, encompassing data and model monitoring, development of test and evaluation metrics, creation of model retraining procedures, actual model updating, retirement and replacement, and model governance to ensure compliance with necessary guidelines and regulations. A specific iteration of this approach is the ML-as-a-Service (MaaS) model, often implemented via application programming interfaces (APIs). Here, contractors oversee the model's entire life cycle, including initial development and ongoing maintenance, while users interact with the model through APIs, typically operating on a pay-per-usage pricing model. This type of model is currently used by companies like OpenAI and by organizations like the XVIIIth Airborne Corps and often works on a pay-per-usage type of pricing scheme.

While the contractor-only approaches present the simplest approach to maintenance planning, they have serious pitfalls that must be considered. For the MaaS model, despite the simplicity of this model, much like any other pay-per-use pricing scheme (e.g., cloud services, SaaS), it can quickly become exorbitantly expensive if there is a lot of use of the service. Additionally, it requires connectivity back to the API to work. So, if the AI-enabled system is meant to work in austere environment or have a lot of usage on the ML-models, going through a MaaS model may be overly costly. Additionally, having contractors perform all the functions of ML maintenance ignores the hard-learned lessons behind the MLOps paradigm; namely the operation of the ML model has been separated from its maintenance and development. A primary reason why MLOps places the development and maintenance of ML models so close to the running of ML models is that these models require constant monitoring and frequent updating (Treveil et al., 2020). In fact, one form of updating, model retraining, can occur as frequently as daily for an ML model in production in an adversarial and dynamic environment. As with our previous computer vision example of detecting objects from a ground perspective, the ML model would need to be, at a minimum, retrained every time the biome changes (e.g., moving from rural to urban) and every time an organization wants to detect a new or different set of objects. Conceivably, such a change in an ML model's operating environment could occur several times over the course of a single operation for a military unit. Thus, given the frequent nature of ML model maintenance, having contractors provide all this maintenance could be cost prohibitive.

At the other end of the spectrum is a service only solution, where servicemembers and DoD civilians are responsible for all of the aforementioned ML model maintenance tasks. While this certainly presents some potential for cost savings in terms of maintenance, the Army and DoD may lack the skill sets in house, in sufficient numbers, to perform some maintenance functions. This is especially true for maintenance functions like designing a test and evaluation scheme for both the ML model and its data as well as determining the right model retraining procedures (e.g., active learning, fine-tuning, using adapters, prompt engineering, etc.). These types of maintenance tasks often take a seasoned data scientist with domain area expertise and, often, advanced education. That said, some of the maintenance tasks actually require very little education and can be learned with suitable training. For example, actually performing model updates, given a guide to the model's retraining procedures, is a trainable task that does not require an advanced educational background. Thus, planning to do the full spectrum of model maintenance in house may be infeasible, given constraints on in house ML expertise.

Recognizing the limitations of both contractor-provided and service-provided maintenance, recent research has advocated for a hybrid approach. This approach leverages contractor expertise while entrusting servicemembers with maintenance execution of the most frequent maintenance tasks (Cruickshank & Kohtz, 2023). Essentially, tasks such as model monitoring and retraining can be trainable for in-house execution within a properly implemented AI-enabled system, allowing servicemembers to handle these responsibilities. Conversely, less frequent, higher-expertise tasks can be outsourced to contractors. The proposed hybrid approach is detailed in Figure 3, illustrating the division of sustainment tasks among the involved components.

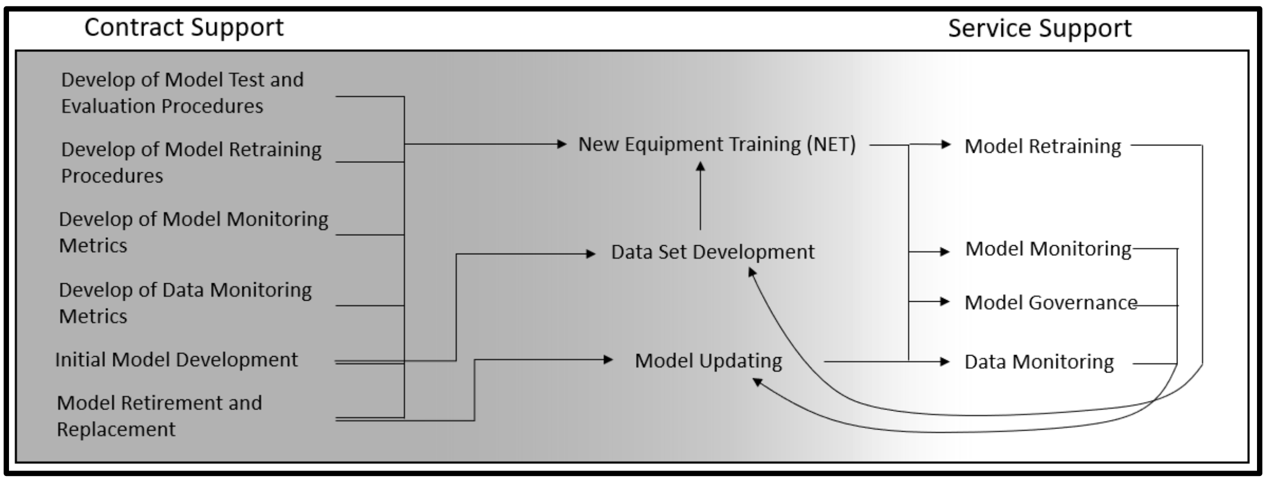


Figure 3. ML Model Sustainment Tasks in a Hybrid Maintenance Plan with Associated Dependencies Between Contractor and Service Maintenance Tasks (Cruickshank & Kohtz, 2023)

Moreover, the hybrid model for ML model maintenance encompasses additional considerations:

- *Data Rights*: Government operators will execute most of the work in the Figure 3 ML model sustainment tasks and maintenance workflow such as model retraining and monitoring. As a result, the negotiation of limited rights between vendors and program offices is plausible and could support our recommended hybrid maintenance plan. The preferred approach with data rights is to negotiate government purpose rights because the funding source will most likely be mixed (government and industry investment) for product deliverables and technical data. A change in a program's future sustainment strategy is still possible with this method and provides program offices with potential sustainment strategy options or adjustments (Cruickshank & Kohtz, 2023).
- *ML Model Touch-Time Analysis*: Since the ML models act as the brain of an AI-enabled system, model retraining is a necessity to enable this paramount function. ML models can require daily, monthly, yearly, or beyond retraining depending on the context (Evidently AI, 2021). A PS BCA must incorporate a retraining requirements analysis. The supplemental analysis should consider how often the data environment changes, whether present in an adversarial environment (i.e., an environment where people developing the data attempt to fool the system by altering the data generation patterns), and the frequency of physical location changes for the data generation process (i.e., a sensor mounted on a platform changes geographic regions). This supplemental analysis will provide a program office a more holistic understanding maintenance requires to accurately inform sustainment cost estimates (Cruickshank & Kohtz, 2023).

A Framework for Estimating Machine Learning Component Maintenance Cost

Maintaining a machine learning model encompasses several task categories, including system monitoring, model updating, data curation, and model curation (Cruickshank & Kohtz, 2023). Although maintaining an ML model involves numerous tasks, these tasks vary in required expertise and frequency of execution (Cruickshank & Kohtz, 2023; Evidently AI, 2021; Zenkevich, n.d.). For instance, retraining an ML model demands more expertise compared to monitoring its usage, and the frequency of model retraining is typically less than that of model monitoring, which should occur whenever the model is in use. Thus, the scope and scale of maintenance for a specific ML-enabled system are determined by variables such as the frequency of maintenance tasks, the expertise needed, and the number of unique ML models requiring maintenance.

Based on these determinants, one can estimate the maintenance costs for ML models within a system using a multiplicative model. Specifically, the maintenance cost estimate for ML models is calculated by multiplying the number of unique ML models in the system by the skill premium per time for each task required to maintain each unique model, further multiplied by the time each skill needs to be performed within a given period. This maintenance cost estimate can be expressed by the following function:

$$\sum_{m=1}^{num\_models} \sum_{i=1}^{num\_tasks} skill\_premium\_per\_time_{m,i} \times \left(\sum_{t=1}^{num\_events} maintenance\_time_{m,i,t} \right)$$

Within this function, the main items to track are the skill premiums and the maintenance times. The skill premiums are the cost (per person, per unit, etc.) to execute a particular task, *i,* for a particular model, *m*. For example, this would be something like the cost per hour to have a maintenance technician retrain an object detection computer vision model. The second major item to track for the cost estimate is the amount of time a particular task is performed for a particular model. Building on the previous example, this would be how many hours it takes that technician to do a retraining of the object detection computer vision model by how many times they do the retraining in a given time period (e.g., per year). With these items, the cost estimate becomes a matter of multiplying the time by the skill premium and summing it up across all time, tasks, and models for an ML-enabled system.

This model does have some important caveats. This model is meant to produce cost estimates at a program-level and for the ML models specifically. Thus, there may be additional costs for specific units as well as for other components of the system, most notably the hardware. Additionally, fixed costs like the training for personnel to execute certain maintenance tasks are not explicitly included in the model but could be easily included in the skill premiums through amortization of the training cost.

A Worked Example

Having now established a cost estimate model for the ML models of a ML-enabled system, we will now work a simple, yet common across the services, example of automated threat recognition (Ferraris, 2021). In this particular case, we will consider an automated threat recognition (ATR) system that is part of a sensor system for a vehicle like a tank or drone. This ATR system will be a simple one consisting of a single, supervised object detection algorithm for one spectra of imagery (e.g., visible, EO, IR, etc.). For this example, we will consider that the ATR system is used at a Brigade Level, within a standard U.S. Army Division. This Brigade has one major training center rotation and 1–2 brigade level exercises that utilize the system per year. It's likely the ATR system would be used in other lower-level training exercises (e.g., Battalion-level training events), but those will most likely have a negligible impact to the ML model maintenance.

Based on the given scenario, there are several model maintenance tasks that would need to be performed each year.

- Once per year, there would need to be a model and software update to keep the ML model on pace with state-of-the-art and address any architectural flaws that may have come out during the previous year. Furthermore, since we are using a supervised ML model, this would also be the chance to revise what classes the model can detect objects of. All of this maintenance will likely need to come with a certification of model safety and ethical usage, given current AI policy guidelines.
- For the training center rotation, there would need to be model retraining for all models and data collection and labeling. This is needed to address the change in scenery that will occur between the host station environment and the training center environment. During the training event, there will also need to be data monitoring and model monitoring, with a possible need for additional model retraining, and data collection and labeling.
- For Brigade-level exercises there would need to be model monitoring and data monitoring. Depending on the nature of the training exercises, there may also be a need for model retraining, and data collection and labeling.

During these training events, the service-provided maintenance would require, at a minimum, two trained officers or NCOs to support operations. In the data management cell, we project an O-3 (Captain) and E-6 (Staff Sergeant) to complete certain maintenance tasks. If a contractor provided this support, two field service representatives (FSR) would provide this support. This maintenance cell would be located at the Brigade Tactical Operations Cell (TOC) to execute the consolidated maintenance of the models across multiple platforms. The maintenance cell would retrieve the periodic inputs of data from Battalion level TOCs and down to the platform (i.e., tank) crew level. The machine learning model data would arrive at the Brigade TOC through the Army provided command and control (C2) networks or from removable hardware through logistic package operations (LOGPAC). Through either method, the transfer of data would occur at least once every 24 hours. If there are some functions that require reach back or depot level support (i.e., model renewal, certification, or data collecting and labeling), the maintenance cell would then transfer the data through similar means to higher echelon support level. Figure 4 is a depiction of the typical machine learning maintenance workflow during Brigade-level collective training or deployments.

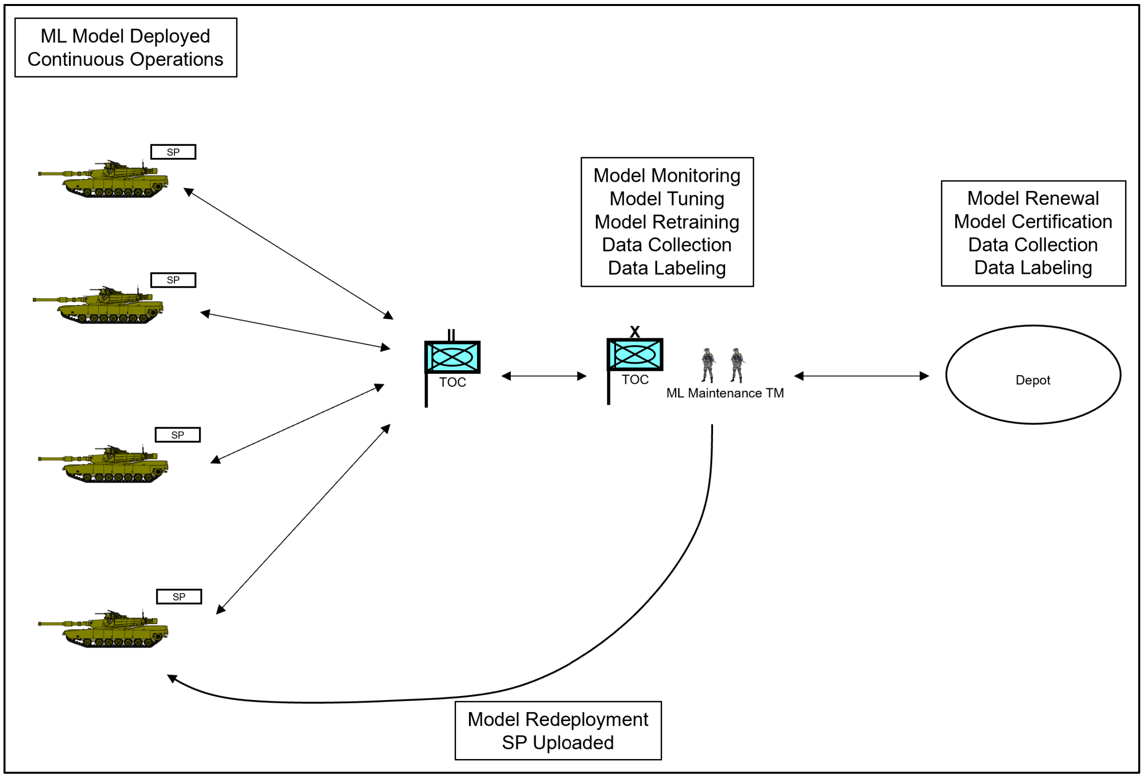


Figure 4. ML Maintenance Operations Workflow

When completing the cost estimate for ML maintenance operations, we used estimates for the skill premium rates. We based the hourly rates for contractor rates on industry median rates. We also created two categories of rates, depending on whether the activity is taking place at home station or in a deployed (i.e., CTC) environment. For the tasks of model retraining, model monitoring, and data monitoring we used the median hourly rate for a data scientist at $60.00 for home station and $80.00 for forward deployed. For the data collection and labeling task we used an industry estimate from labeling services of $40.00 per hour for home station and $75.00 for forward deployed. And, for the model renewal and certification, which includes all of the tasks of updating the model to a new model, instruction in that model's monitoring metrics and retraining procedures, and implementation testing of the new model we used a senior or chief data scientist hourly rate of $120.00 an hour. For the service-provided rates, we used the median hourly salary for an O-3 (Captain) for model retraining and the median hourly salary for an E-6 (Staff Sergeant) for model monitoring, data monitoring, and data collection and labeling, based on who is most likely to perform the particular skills. We also included in an additional $5.00 per hour and $4.00 per hour, respectively, for amortized training that these service-provided personnel would need to execute the maintenance tasks. We also observe that contractor rates will almost certainly be higher than service-provided rates for these tasks and that some tasks, due to their high skill requirements may only be able to be done by contractors (i.e., model renewal and certification).

Similar to the skill premium rates, we also used estimates for the number of hours per skill execution. We allowed for differences in hours depending on the event. For example, a CTC rotation will require more hours for some skills than home station training or periodic maintenance. For model retraining, we estimate an average of 32 hours per retraining, based on two people doing, on average 16 hours of work. We also include the time it takes to put the model back into operation following retraining as part of this estimate. For data collection and labeling, we estimated around three people doing 40 hours of work for an average of 120 hours. For both data and model monitoring we estimated around 40 hours of work for two people for an average of 80 hours per event. Finally, for the model certification and renewal we estimated around 180 hours to accomplish this, but note this estimate could vary wildly as the tasks that comprise this task have not been done all together around an ML-enabled system yet. Each of these estimates were increased for a CTC rotation, with mean hours for model retraining, data collection and labeling, data monitoring, and model monitoring going to 40, 320, 160, and 160 hours respectively. Finally, it's important to note that these estimates are not meant to be definitive, but rather meant to illustrate how there is a cost to ML-maintenance and that this cost can vary significantly between different maintenance strategies.

Based on the scenario previously outlined and the estimates for the drivers of ML maintenance, we propose two different maintenance strategies. The first strategy is to use contractors for all ML model maintenance. The second is to use service provided maintenance for all tasks except for those tasks under model renewal and certification and allow for some additional contractor support on data collection and labeling.

Once we estimated the touch time requirements for maintenance of ML models, the cost estimates were expanded to encompass a typical 20-year service life for a ground platform (DoD, 2020). The cost estimates we developed for the potential maintenance strategies rely on the current year (2024) to normalize costs. As a result, the approach utilizes the current-year cost approach because current-year estimates are necessary when conducting comparisons (DoD, 2021). The current-year costs, with 2024 as the base year, utilizes the Army/Navy Joint Inflation Calculator Indices (JIC) to support the comparison of product support strategies (Army Financial Management & Comptroller, 2023). The first strategy with contractor only support utilizes the Operation and Maintenance Army (OMA) appropriation weighted index because program offices must obligate funds in the fiscal year; however, disbursements from the treasury could happen after the initial obligation year (DoD, 2021). The second strategy that utilizes a hybrid approach for product support incorporated a composite index for Military Pay Army (MPA) appropriation to capture a blend of inflation and escalation cost increases for military pay since this option utilizes service members (DoD, 2021). The weighted composite index for this option is applied to cumulative costs for service members and contractor estimates because there could be OMA fund disbursements after the obligation year for contractors and our model includes the assumption that the pay escalation will be similar between contractor and military personnel. These cost estimate inputs are necessary to support a fair and realistic comparison between the two options, and the model in current-year dollars provides the best picture for decision makers to understand the most cost-effective option.

A system with a 20-year service life will have a phase-in, steady-state, and phase-out periods (DoD, 2020). The base year for our notional ATR model starts in 2024 and ends after a 20-year service life in 2044. The following graph (Figure 5) displays the cumulative cost estimates of these two strategies for just one armored brigade over the typical 20-year service life period.

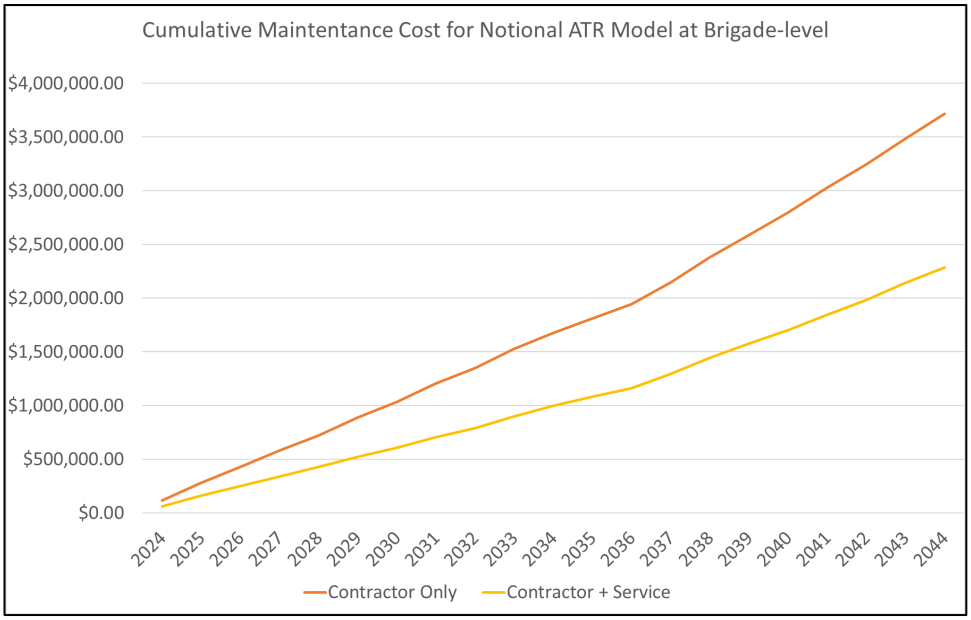


Figure 5. Cumulative Maintenance Cost for Notional ATR Model at Brigade-Level

From this plot we can observe that there is a significant cost per year for each strategy, and that a contractor only strategy grows in cost more quickly over time than contractor + service provided maintenance strategy. The graph shows the cost implications of adding one ATR system to vehicles for just one armor brigade in the entire Army. The difference in ML model maintenance between both options exceeds $1,000,000 across the service life for one brigade. As a result, the costs for the ATR Notional system across all armor brigades will have a larger impact. The plot in Figure 6 depicts the costs associated with implementing the ATR Notional system across all nine armor brigades in the Army.

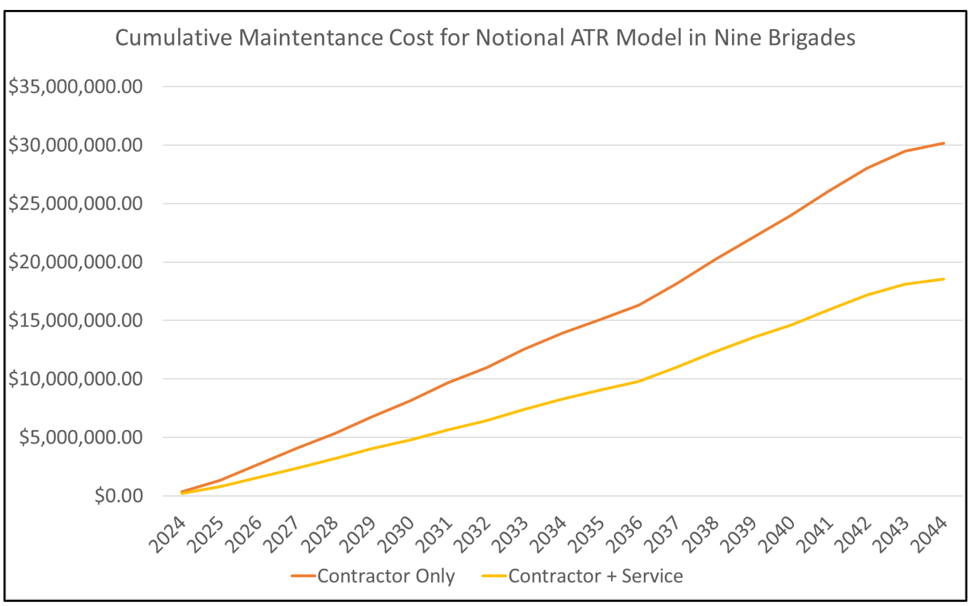


Figure 6. Cumulative Maintenance Cost for Notional ATR Model in Nine Brigades

The cost estimation of a program of record (POR) for the Notional ATR Model in a Sensor Payload highlights the importance of a hybrid machine learning maintenance strategy. The indices and use of 2024 as the current-year funds in the ATR Notional System cost estimate for one brigade is the same for a POR supporting nine brigades and across all of this research's cost estimates. The POR model incorporated typical planning considerations for the operations and sustainment phase of a program of record. The cost estimation model has a phase-in period of two years (2024 and 2025) with three and then six brigades utilizing the system for the first two years. Steady-state operations last 16 years from 2026 to 2042. The phase-out period reduces the number of systems for six and then three brigades at the end of a 20-year service life in 2044. Once a system is at scale and fielded across nine armor brigades, the difference in the cost effectiveness of maintenance strategies become more apparent. For example, the potential cost savings with a hybrid approach implementation is nearly $11,595,000 with the contractor only strategy estimate close to $30,158,000. The difference in strategy maintenance costs demonstrates how imperative it is for the services to support with an appropriate level of organic support for AI-enabled systems.

The next graph depicts how maintenance of a AI-enabled system will be different than typical hardware and software maintenance of a system. The amount of touch time required for ML systems could potentially surpass the costs of a hardware and software stack. In order to understand how ML model maintenance may compare to a system maintenance plan, we utilized an analogous estimating approach from two different sources to establish a notional hardware and software stack for the Sensor Payload with the ATR model. First, for software, the cost model utilized a similar system (Common Sensor Payload) cost estimate from a previous business case analysis for software sustainment. The previous analysis evaluated multiple support strategies, and we utilized the estimate with a 50/50 maintenance ratio of contractor and government support (Software Engineering Center [SEC], 2022). This software sustainment strategy incorporates software, firmware, and obsolescence updates to the fielded systems. Second, for the hardware maintenance, the cost model incorporated the example component level cost estimate for a avionics or electronic subsystem from the Operating and Support Cost-Estimating Guide (DoD, 2020). This example of an O&S hardware cost estimate is for a notional system like the common sensor payload that incorporates spares, parts, labor, and recurring training to support hardware maintenance for 10 units with host platforms (DoD, 2020). We then scaled the software and hardware cost estimates to model costs across nine brigades for a 20-year service life in current-year dollars (2024). The cost model approach illustrates how ML model maintenance occurs in addition to the hardware and software stack maintenance of a system. As a result, given the potential for extensive touch time, the ML model maintenance will increase existing maintenance and could potentially be higher than the current paradigm of system software maintenance while less than hardware maintenance. Figure 7 shows how machine learning model maintenance may compare to existing system maintenance.

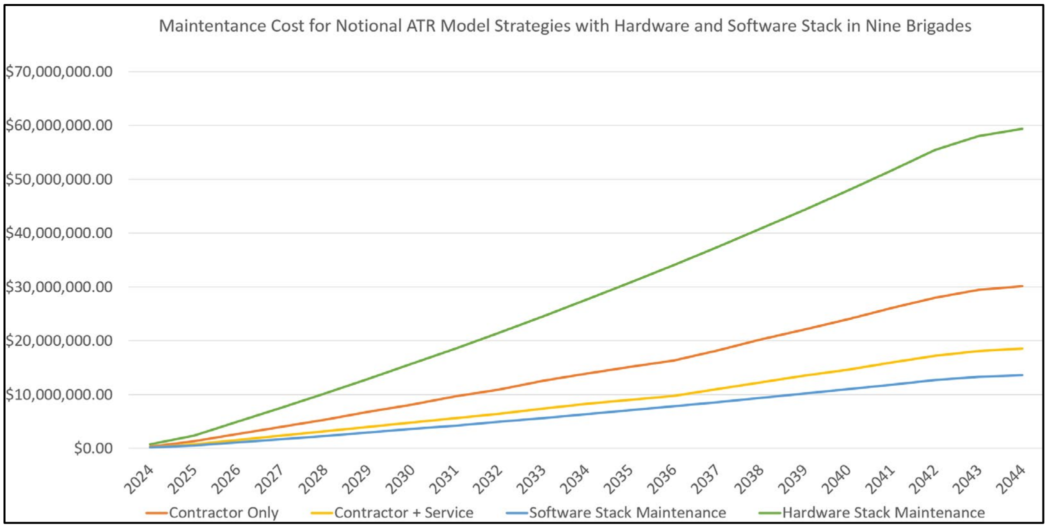


Figure 7: Cumulative Maintenance Cost for Notional ATR Model Strategies with Hardware and Software Maintenance in Nine Brigades

Conclusion

Implementation of AI and ML models to provide fighting forces at the tactical edge remains a priority to gain the advantage on the battlefield. The services and program offices chartered with delivering capability to the warfighter must understand the nuances of ML model maintenance when developing product support strategies. A mindset of treating AI and ML capability as a system within a system is imperative when developing strategies and estimating costs. Machine learning models are an added layer of maintenance on top of a system's typical hardware and software maintenance, and program offices must treat it differently. The graph in Figure 8 highlights how adding or implementing ML capability in a program of record will increase the O&S costs of a program of record. This is why relying on organic support for some maintenance support is a must do when planning sustainment strategies.

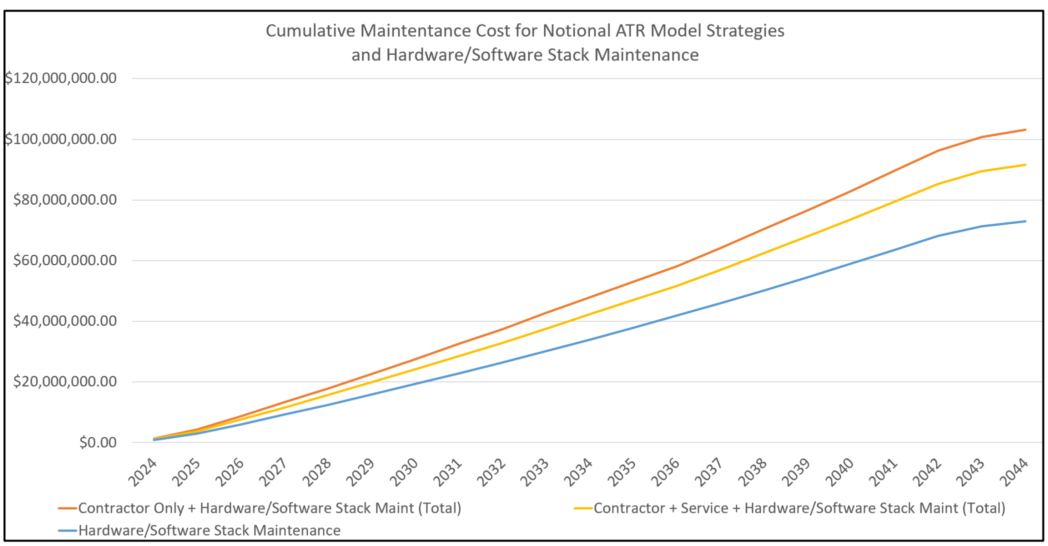


Figure 8: Cumulative Maintenance Cost for Notional ATR Model Strategies Plus Hardware/Software Stack Maintenance

The amount of touch time required for ML enabled systems is the main cost driver for these systems. As a result, services should rely on some level of organic support to maintain the models and rely on industry for the more technically advanced tasks as highlighted in the background section of this research. A contractor only approach will become cost prohibitive when implementing this new technology across the entire fleet of sensors or systems in all of the branches of service. Our evaluation of the contractor only approach and a hybrid approach of contractor and service support strategies is a framework for program offices to utilize when implementing machine learning model capabilities. Incorporation of these methods for AI-enabled systems early in the planning process will pay dividends later in the service life of programs.

Finally, its important to note that recent innovations in the ML space in generative models, like Large Language Models and Vision Language Models, can alter this maintenance landscape. These models are applicable to a wider array of tasks, without any additional training of the models, but require new forms of interaction from the users with techniques like prompt engineering. As such, these models could alter the maintenance landscape by making certain types of maintenance tasks, like fine-tuning occur almost in stride with the operation of the model which further necessitates that maintenance of the model be capable of being conducted within the services. We leave investigation of the maintenance of these newer generative types of AI to future work.

The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

References

Army Financial Management & Comptroller. (2023). *Joint inflation calculator*. https://www.asafm.army.mil/Cost-Materials/Cost-Models/#rates-guidance

Breck et al., (2017) *The MLTest score: A rubric for ML production readiness and technical debt reduction*. https://storage.googleapis.com/pub-tools-public-publication-data/pdf/aad9f93b86b7addfea4c419b9100c6cdd26cacea.pdf

Camm, F., Whitmore, T. C., Weichenberg, G., Sheng, T. L., Carter, P., Dougherty, B., Nalette, K., Bohman, A., & Shostak, M. (2021). *Data rights relevant to weapon systems in Air Force special operations command* (Report No. RR-4298-AF). RAND. https://www.rand.org/pubs/research_reports/RR4298.html

Cruickshank, I., & Kohtz, S. (2023). Acquiring maintainable AI-enabled systems. *Excerpt from the Proceedings of the Twentieth Annual Acquisition Research Symposium,* 59–67. https://dair.nps.edu/handle/123456789/4852

Department of Defense. (2014). *DoD product support business case analysis guidebook.* https://www.dau.edu/tools/Lists/DAUTools/Attachments/127/Product-Support-Business-Case-Analysis-(BCA)-Guidebook.pdf

Department of Defense. (2016). *Operating and support cost management guidebook.* https://www.dau.edu/tools/Lists/DAUTools/Attachments/126/Operating-and-Support-Cost-Management-Guidebook.pdf

Department of Defense. (2020). *Operating and support cost-estimating guide*. https://www.cape.osd.mil/files/OS_Guide_Sept_2020.pdf

Department of Defense. (2021). *Inflation and escalation best practices for cost analysis (analyst handbook).* https://www.cape.osd.mil/files/Reports/OSDCAPEEscalationHandbook2021.pdf

Department of Defense. (2022a). *Defense budget overview: United States Department of Defense fiscal year 2023 budget request.* https://comptroller.defense.gov/Portals/45/Documents/defbudget/FY2023/FY2023_Budget_Request_Overview_Book.pdf

Department of Defense. (2022b). *Product support manager guidebook.* https://www.dau.edu/tools/Lists/DAUTools/Attachments/129/Product-Support-Manager-(PSM)-Guidebook.pdf

Evidently AI. (2021, July 01). *When to retrain a machine learning model? Run these 5 checks to decide on the schedule.* https://www.kdnuggets.com/2021/07/retrain-machine-learning-model-5-checks-decide-schedule.html

Ferraris, P. (2020, February 05). *Aided detection on the future battlefield*. https://www.army.mil/article/232074/aided_detection_on_the_future_battlefield.

General Services Administration. (2023, March 1). *Defense federal acquisition regulation supplement (DFARS) 227.7103-5 government rights.* https://www.acquisition.gov/dfars/227.7103-5-government-rights

Guariniello, C., Balasubramani, P., & DeLaurentis, D. (2021). A system-of-systems approach to enterprise analytics design: Acquisition support in the age of machine learning and artificial intelligence. *Proceedings of the Nineteenth Annual Acquisition Research Symposium*, 205–217. https://dair.nps.edu/handle/123456789/4519

Moore, A., Hebert, M., & Shaneman, S. (2018). The AI stack: A blueprint for developing and deploying artificial intelligence. *Proceedings of SPIE Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IX*. https://www.ri.cmu.edu/publications/the-ai-stack-a-blueprint-for-developing-and-deploying-artificial-intelligence/

Nagy, B. (2022). Tips for CDRLs/requirements when acquiring/developing AI-enabled systems. *Proceedings of the Nineteenth Annual Acquisition Research Symposium*, 218–241. https://dair.nps.edu/handle/123456789/4587

National Security Commission on Artificial Intelligence. (2021). *Final report*. https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf.

Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD[A&S]). (2021, November 4). *Product support management for the adaptive acquisition framework* (Department of Defense Directive 5000.91). Department of Defense. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500091p.PDF?ver=qk1slCU3Y0c1acIDocWyJA%3d%3d

Patruno, L. (2019, June 10). *The ultimate guide to model retraining.* https://mlinproduction.com/model-retraining/#:~:text=Quickly%20changing%20training%20sets%20might,require%20monthly%20or%20annual%20retraining

Schinasi, K. V. (2003). *Best practices: Setting requirements differently could reduce weapon systems' total ownership costs.* Government Accountability Office.

Sculley et al. (2014). *Machine learning: The high interest credit card of technical debt*. https://research.google/pubs/machine-learning-the-high-interest-credit-card-of-technical-debt/

Sculley et al. (2015). *Hidden technical debt in machine learning systems.* https://papers.nips.cc/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf

Software Engineering Center (SEC). (2022). *Common sensor payload (CSP) product support business case analysis (PS BCA) for software sustainment*. Communications-Electronic Command.

Talby, D. (2018, June 5). *Lessons learned turning machine learning models into real products and services.* https://www.oreilly.com/radar: https://www.oreilly.com/radar/lessons-learned-turning-machine-learning-models-into-real-products-and-services/

Treveil, M., Omont, N., Stenac, C., Lefevre, K., Phan, D., Zentici, J., . . . Heidmann, L. (2020). *Introducing MLOps.* O'Reilly.

Visengeriyeva, L., Kammer, A., Bär, I., Kniesz, A., Plöd, M., & Eberstaller, S. (2023, March 30). *MLOps principles*. https://ml-ops.org/content/mlops-principles

Zinkevich, M. (n.d.). *Rules of machine learning: Best practices for ML engineering.* https://developers.google.com/machine-learning/guides/rules-of-ml/

System Acquisition Cost Modeling Initiative to Quantify AI Assistance

Ray Madachy—is a research staff member at the Institute for Defense Analyses (IDA), with expertise in legal, regulatory, and financial matters relating to the DoD. Prior to joining IDA in 2005, he held a number of positions in the commercial sector. He holds a bachelor's degree in economics from Dickinson College; a Doctor of Jurisprudence degree from Georgetown University Law Center; and a Master of Business Administration degree from The University of Chicago Booth School of Business. [rjmadach@nps.edu]

Ryan Bell—is a research staff member at the Institute for Defense Analyses (IDA) specializing in contract matters relating to the DoD. She has a doctoral degree in economics from The Ohio State University. [ryan.bell@nps.edu]

Ryan Longshore—is a research staff member at the Institute for Defense Analyses (IDA) specializing in contract matters relating to the DoD. She has a doctoral degree in economics from The Ohio State University. [ryan.longshore@nps.edu]

Abstract

Artificial Intelligence (AI) based tools that assist in generating system artifacts are transforming systems and software engineering lifecycles. Drastic reductions in effort are possible using tools that use large language models (LLMs). This research addresses the new challenges in systems and software cost modeling with the introduction of cost factors and size measures to incorporate into existing parametric cost models.

Keywords. Systems Acquisition Cost Modeling, Parametric Cost Modeling, Artificial Intelligence, Generative Artificial Intelligence, Large Language Models (LLMs), Systems Engineering, Software Engineering, COSYSMO, COCOMO

Introduction and Background

A disruptive transformation is occurring on how the Navy and the rest of the DoD develops, delivers, and sustains systems due to AI assistance in the systems and software engineering processes. AI tools can support virtually all non-hardware production lifecycle aspects from concept, AoA, architecture, requirements, design, software, V&V, testing, etc. A research goal is to better understand and codify the advantages and pitfalls of integrating AI into systems and software processes. We consider the benefits, challenges, dangers of over-reliance and potential inefficiencies.

We have observed that drastic reductions in effort are possible using AI assistant tools that use large language models (LLMs). LLMs are a type of generative AI that utilize a deep learning algorithm to generate human-like text based on natural language prompts. One typically interfaces with a chatbot such as ChatGPT, Gemini, Claude, Copilot and many others. They are well suited for tasks such as language translation, text summarization, and question answering. Some LLMs are exceptionally good at generating code and text-based system models like SysML 2.

AI-based tools can assist in generating system artifacts across virtually all phases and activities. The research initiative examines the quantitative AI impacts by lifecycle phase and activity since their effects may vary greatly. Examples of potential cost benefits and risks for traditional technical processes in systems and software engineering are shown in Tables 1 and 2.

Table 1. Example Engineering Activity Cost Benefits

Activity	Benefits
Requirements	AI tools can help clarify terminologies and concepts on-the-fly, reducing the need for prolonged meetings or external consultations. Developers or analysts can query AI tools for insights or comparative analysis, which might help in refining requirements.
Design	Engineers can quickly consult AI tools for recommended design patterns or architectural practices suitable for their problem. Preliminary design ideas can be discussed with AI tools for quick feedback.
Code	Developers can seek assistance on coding challenges, syntax, and algorithmic solutions. AI tools can assist in code reviews, highlighting potential pitfalls or anti-patterns.
Testing and Integration	AI tools can suggest potential edge cases or testing scenarios. For failing tests or integration issues, developers can discuss potential causes and solutions with AI tools.
Maintenance	AI tools can assist in understanding old codebases, suggesting potential refactoring techniques, or identifying deprecated methods. For known errors or bugs, AI tools can suggest common solutions or workarounds.

Table 2. Example Engineering Activity Cost Risks

Activity	Risks
Requirements	Relying too much on AI tools for domain-specific knowledge might lead to missed nuances that an expert in the field would be aware of.
Design	Over-relying on AI for design decisions without human review can lead to suboptimal choices.
Code	If developers use AI tool suggestions verbatim without understanding, it might introduce bugs or inefficient code. Waiting on AI tool responses for every small issue can become a crutch and delay development if developers stop trying to problem-solve on their own.
Testing and Integration	If AI tool suggestions are taken without thorough review, it might lead to unnecessary tests or efforts spent on non-issues. If teams over-rely on AI tools for maintenance, they might overlook deeper architectural or design issues that require human expertise.
Maintenance	AI tools can assist in understanding old codebases, suggesting potential refactoring techniques or identifying deprecated methods. For known errors or bugs, AI tools can suggest common solutions or workarounds.

This research is using the comprehensive IS0 15288 lifecycle standard for systems and software engineering (International Organization for Standardization [ISO], 2015) as a framework for data collection and analysis of detailed activities. The activities in Tables 1 and 2 contain a subset of the technical processes in ISO 15288. While this initiative is first addressing software development cost with formal definitions and data collection, we are performing allied research in systems engineering AI assistance and cost impacts.

This research is quantifying the impact against standard systems and software engineering tasks, generating and analyzing artifacts with the assistance of AI. Benchmarks are first necessary for well-defined engineering tasks for quantification and reproducibility.

SysEngBench is being designed to evaluate LLMs in the context of systems engineering concepts and applications to support benchmarking (Bell et al., in press). It will encompass a comprehensive set of tasks derived from core systems engineering processes, including requirements analysis, system architecture design, risk management, and stakeholder communication. By leveraging a diverse array of real-world and synthetically generated scenarios, *SysEngBench* aims to provide an assessment of LLMs' ability to interpret complex engineering problems and generate innovative solutions.

Our research also explores the ability of current LLMs to generate, modify, and query Systems Modeling Language (SysML) v2 models (Longshore et al., in press). Techniques such as Retrieval-Augmented Generation (RAG) are utilized to add domain-specific knowledge to an LLM and improve model accuracy. Preliminary case studies indicate that the number of prompts to generate models can be minimized.

Method

This research addresses the new challenges in systems and software cost modeling with model definitions, a lifecycle standard, and data analysis process to calibrate the model. This includes new cost factors and size measures to incorporate into existing parametric cost models. Empirical data collection and analysis for model calibration is also underway.

Already, there is very strong convincing data that substantial labor can be saved in steady-state AI tool usage by individuals and teams. To address the cost impacts, we have developed a road map for advancing the cost models by leveraging existing modeling and measurement frameworks. We are using the Constructive Cost Model (COCOMO) framework and calibration procedures (Boehm, 1981; Boehm et al., 2000).

The cost modeling and measurement framework incorporates a new factor for "AI Assistance Usage" with a defined rating scale and data analysis process to calibrate it. An online data collection and Delphi survey to improve the model with expert judgment has been developed for the community. A new measure, "query points," is being refined to quantify the size and complexity of the AI generated solutions.

From systems and software engineering case studies, we are gathering empirical data on generated solution sizes, actual effort, and effort estimates without AI assistance. Subsequent case studies will address larger scale team and enterprise processes assisted with AI.

Parametric Modeling

The general effort formula used in the parametric systems and software cost models is:

$$Effort = A * Size^{B} * \prod_{i=1}^{N} EM_i$$

where

- *Effort* is in Person-Months (PM)
- *A* is a constant derived from historical project data
- *Size* is a measure of the work product

- *B* is an exponent for the diseconomy of scale
- EM_i is an effort multiplier for the i^{th} cost driver. The geometric product of *N* multipliers is an overall Effort Adjustment Factor (EAF) to the nominal effort.

Size is interpreted within the context of the specific work being estimated in the units of the work products. The effort multipliers cover factors for product, platform, personnel and project attributes that affect cost. The resulting top-level effort can be decomposed for each phase, activity or increment.

Example Cost Driver and Effort Multipliers

An example cost driver, *Applications Experience*, from COCOMO II is visualized in Figure 1 with its effort multipliers. The effort multipliers for each rating represent the relative effort to Nominal. For example, the EM for a Very Low rating of Applications Experience is 1.22, indicating a 22% increase in effort from Nominal. The Nominal rating is always 1.0 by definition for a typical project. The High rating EM is 0.88, or a 12% decrease in effort from Nominal. The overall Effort Multiplier Ratio (EMR) for Applications Experience is the ratio of the highest to lowest multipliers, or 1.22/.81 = 1.5.

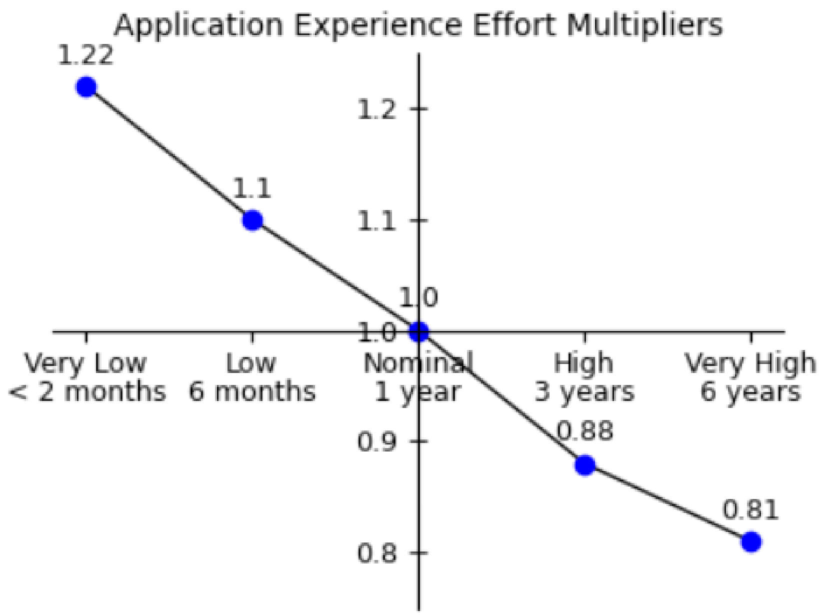


Figure 1. Application Experience Effort Multiplier Example

The initial rating scale for the proposed cost driver "AI Assistance Usage" has been defined using the COCOMO framework. It consists of five ratings from Very Low to Very High, corresponding to the degree of AI usage on a project per Table 3. The default setting is Nominal, corresponding to a typical project. The data collection will be used to calibrate the effort multipliers for each rating level.

Table 3. AI Assistance Usage Initial Rating Scale

Very low	Low	Nominal	High	Very High
Minimal to no AI assistance. Development relies primarily on traditional methods and tools. AI tools may be present but are rarely, if ever, consulted.	Occasional AI consultation, typically for clarification or basic information retrieval. AI tools are not deeply integrated into the development workflow.	Regular use of AI tools for various tasks like code help, design insights, or testing assistance. AI tools are a recognized part of the toolkit but aren't central to development	Frequent and strategic use of AI assistance. AI tools play a central role in multiple phases of development, from design to code review.	AI tools are deeply ingrained in most development phases. They are crucial for decision making, problem solving, and automating specific tasks. The development process is designed around maximizing AI tool benefits.

We have also identified other affected cost factors and parameters for using generative AI. For example, the relative cost of achieving reliability may change, and AI may help reduce impacts of experience and capability. Overall cost model coefficients will change. Usage of AI will shortly become an assumed skillset of engineers. Subsequent data collection will help us assess these impacts as well.

The initial definition is oriented to software. The factor definition and its data collection are setting the stage for further exploration into systems engineering process impacts. We are defining an analogous usage factor for systems engineering to incorporate in the Constructive System Engineering (COSYSMO) model (Valerdi, 2005). In our research, we are generating SysML 2 artifacts and capturing data on effort, solution accuracy, size and complexity for activities covered in COSYSMO. In additional case studies, we will collect similar data from large team projects.

We are also investigating phase sensitive effort multipliers to account for different AI tool impacts across the lifecycle. We are codifying the practices by phase and activity, and empirical data collection is being aligned with those in the ISO/IEC/IEEE 15288 lifecycle.

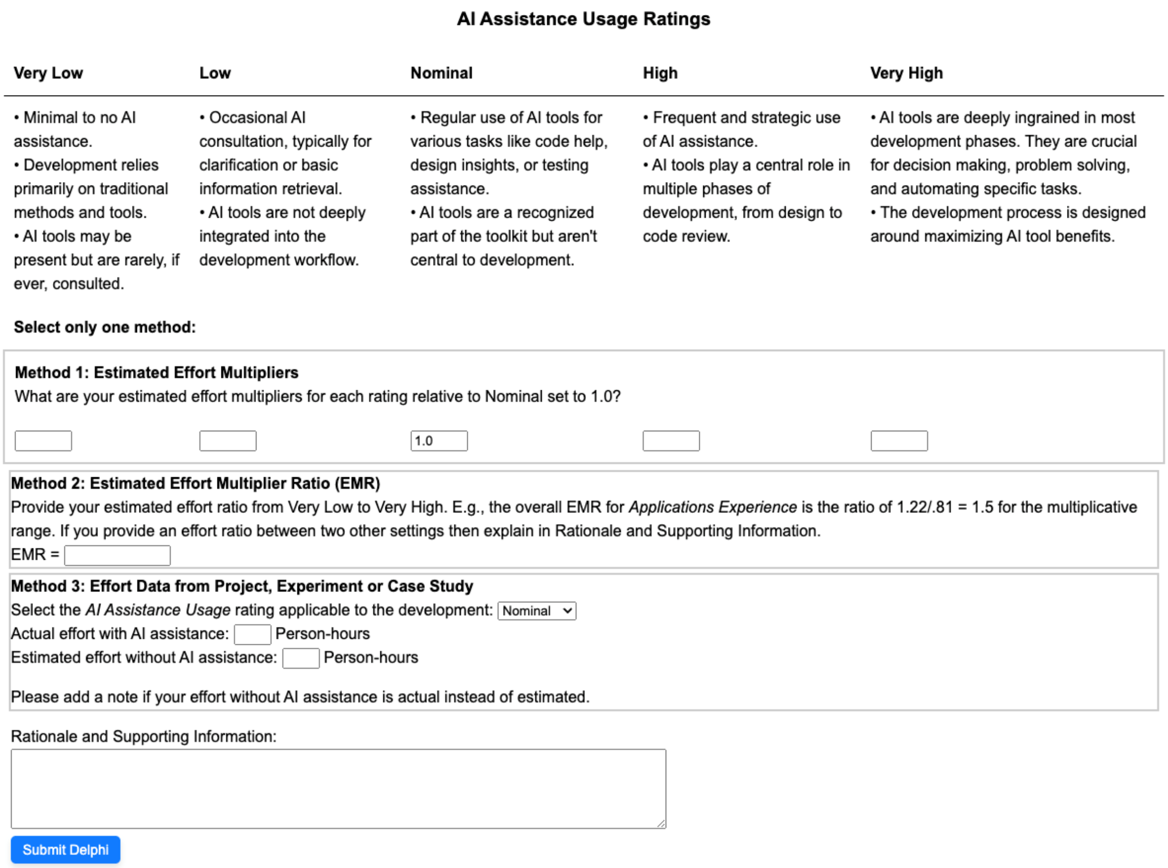
Data Collection and Analysis

Different empirical methods are used for parametric cost model analysis and calibration. These include:

- multi-project data collection in conjunction with other cost factors (e.g., COCOMO II; Boehm et al., 2000)
- controlled group experiments (e.g., Github Copilot; GitHub, 2022)
- Delphi surveys for expert judgment
- Bayesian approaches combining empirical project data and Delphi results (e.g., COCOMO II; Boehm et al., 2000; Chulani et al., 1999)
- small-scale empirical case studies and expert judgment

A variety of data sources are being drawn from to support model calibration and provide insights. Multi-project data collection in conjunction with other cost factors is going forward for the COCOMO III model. Small-scale empirical case studies and controlled group experiments are being performed. We are also collecting classroom data.

We have developed an online Delphi survey form to capture both expert judgment and actual data to help calibrate the model. A portion is shown in Figure 2. It is available at http://softwarecost.org/data/ai. The data collection is being supported by the Boehm Center for Systems and Software Engineering.



AI Assistance Usage Ratings

Very Low	Low	Nominal	High	Very High
• Minimal to no AI assistance. • Development relies primarily on traditional methods and tools. • AI tools may be present but are rarely, if ever, consulted.	• Occasional AI consultation, typically for clarification or basic information retrieval. • AI tools are not deeply integrated into the development workflow.	• Regular use of AI tools for various tasks like code help, design insights, or testing assistance. • AI tools are a recognized part of the toolkit but aren't central to development.	• Frequent and strategic use of AI assistance. • AI tools play a central role in multiple phases of development, from design to code review.	• AI tools are deeply ingrained in most development phases. They are crucial for decision making, problem solving, and automating specific tasks. • The development process is designed around maximizing AI tool benefits.

Select only one method:

Method 1: Estimated Effort Multipliers
What are your estimated effort multipliers for each rating relative to Nominal set to 1.0?

1.0

Method 2: Estimated Effort Multiplier Ratio (EMR)
Provide your estimated effort ratio from Very Low to Very High. E.g., the overall EMR for *Applications Experience* is the ratio of 1.22/.81 = 1.5 for the multiplicative range. If you provide an effort ratio between two other settings then explain in Rationale and Supporting Information.
EMR =

Method 3: Effort Data from Project, Experiment or Case Study
Select the *AI Assistance Usage* rating applicable to the development: Nominal
Actual effort with AI assistance: Person-hours
Estimated effort without AI assistance: Person-hours

Please add a note if your effort without AI assistance is actual instead of estimated.

Rationale and Supporting Information:

Submit Delphi

Figure 2. Data Collection Form Portion

Ideal Effort Multiplier

A method is needed to isolate the effect of AI assistance among the contaminating effects of other parametric cost factors when analyzing multiple project datapoints. The Ideal Effort Multiplier (IEM) method will be used to determine calibrated multipliers for each project and perform regression across the rating scale to attain global effort multipliers for the model. The IEM quantifies the contribution of AI Assistance Usage eliminating other cost factor sources per:

$$IEM(P, \text{Cost Factor}) = PM(P, \text{actual}) / PM(P, \text{Cost Factor})$$

where

- *IEM(P*, Cost Factor) is the ideal effort multiplier for project *P*
- *PM* is Person-Months of effort
- *PM(P*, actual) is the actual development effort of project *P*
- *PM(P*, Cost Factor) is the cost model estimate excluding the Cost Factor

The IEM method is visualized in Figure 3 with representative values from COCOMO II. The values at each setting represent the best fit against all the project data points.

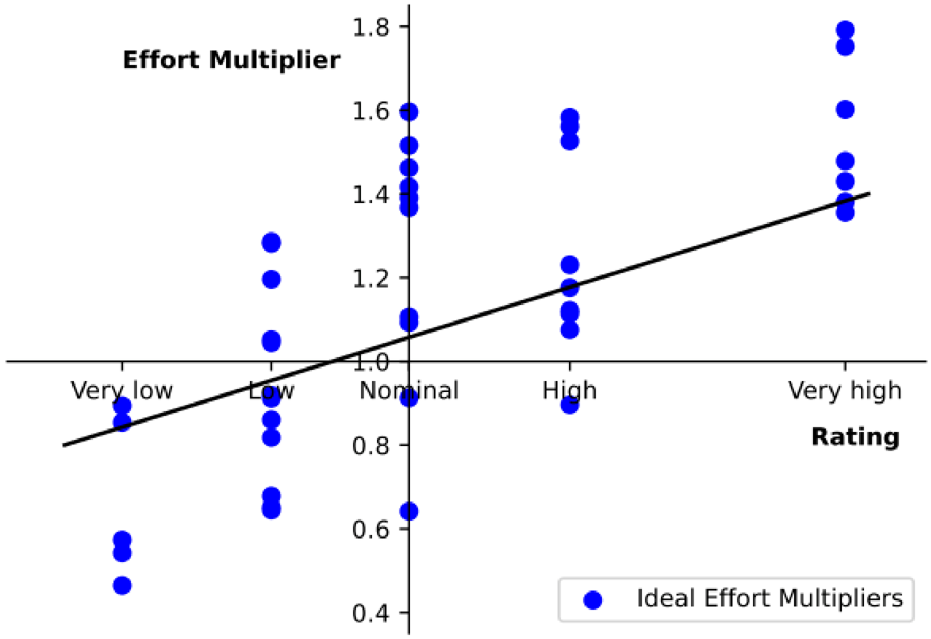


Figure 3. Ideal Effort Multiplier Normalization Method

Detailed Phase Impacts

Phase-sensitive effort multipliers account for different impacts across the lifecycle versus the project level effects in the previous multipliers. This is necessary because some cost drivers vary more across phases than others. An example is shown in Figure 4 from the Detailed COCOMO model for the factor *Use of Software Tools* (Boehm, 1981). It is a highly relevant example because software tools now include AI assistance.

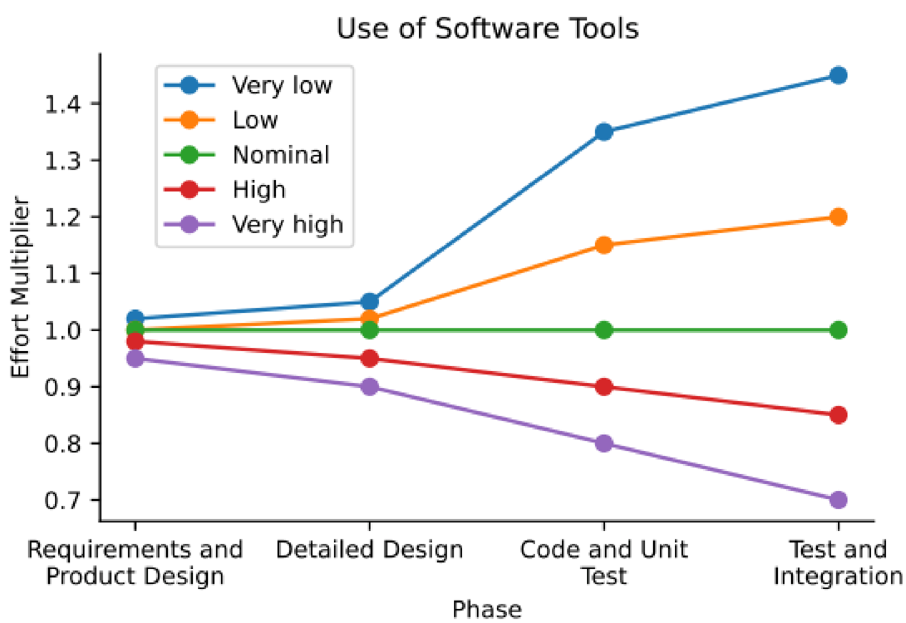


Figure 4. Phase Sensitive Effort Multiplier Example

Empirical data collection will be commensurate with phases so that AI impact will be evaluated by both phases and activities. This will include codifying the practices and ratings with detailed descriptions and examples.

Current and Future Work

We are early in the initiative, and the community is highly encouraged to provide feedback on the model definitions and submit data and feedback on the data collection. Community support is imperative to develop the new models. We are instituting the Delphi data collection and will continue iterative analysis with COCOMO III research to update the models. For this, we will provide open-source cost modeling tools with new factor(s) in the models for public usage.

A current research focus is on how query task complexity impacts AI correctness and effort impact. We are elaborating query points as a complexity measure for this to measure SysML 2 model artifacts. We will perform further analysis of AI tool impacts across lifecycle aligning artifacts and effort data with ISO/IEC/IEEE 15288 systems and software engineering phases and activities. This harmonization will also help address large scale team and enterprise processes assisted with AI.

References

Bell, R., Madachy, R., & Longshore, R. (in press). Introducing SysEngBench: A novel benchmark for assessing large language models in systems engineering. *Proceedings of the 2024 Acquisition Research Symposium*.

Boehm, B., Abts, C., Brown, W., Chulani, S., Clark, B., Horowitz, E. … Steece, B. (2000). *Software cost estimation with COCOMO II*. Prentice-Hall.

Chulani, S., Boehm, B., & Steece, B. (1999, December). Bayesian analysis of empirical software engineering cost models. In *IEEE Transactions on Software Engineering*, *25*(4).

GitHub. (2022). *Research: Quantifying GitHub Copilot's impact on developer productivity and happiness*.

https://github.blog/2022-09-07-researchquantifying-github-copilots-impact-on-developerproductivity-and-happiness.

International Organization for Standardization. (2015). *ISO/IEC 15288:2015 Systems engineering – System life cycle processes*. Tech. Rep. International Organization for Standardization.

Longshore, R., Madachy, R., & Bell, R. (in press). Leveraging generative AI to create, modify, and query MBSE models. *Proceedings of the 2024 Acquisition Research Symposium*.

Valerdi, R. (2005). *The constructive systems engineering cost model (COSYSMO)* [Doctoral dissertation, University of Southern California].

PANEL 24. INNOVATIVE IDEAS AND INSIGHTS FOR IMPROVING PROGRAM RESOURCING ACROSS SEAMS

Thursday, May 9, 2024	
3:45 p.m. – 5:00 p.m.	**Chair: Elizabeth Bieri,** Director of Research for the Commission on Planning, Programming, Budgeting and Execution (PPBE) Reform ***Innovative Ideas and Insights for Improving Program Resourcing across Seams*** Edward Cardon, AIRC ***Research and Development: DOD Benefited from Financial Flexibilities but Could Do More to Maximize Their Use*** Leslie Ashton, GAO ***PPBE, Technology Transition, and "The Valley of Death"*** Olivia Letts, GMU

Elizabeth Bieri—is the Director of Research for the Commission on Planning, Programming, Budgeting and Execution (PPBE) Reform. Ms. Bieri has spent her career in financial management, logistics, and as a cost and price analyst in the national defense space. Prior to this position, she was the Deputy Director for Budget in the Office of the Assistant Secretary of Defense for Special Operations and Low-Intensity Conflict, Secretariat for Special Operations. Ms. Bieri holds a Master of Arts in National Security and Strategic Studies from the College of Naval Warfare at the Navy War College and a Bachelor of Arts from Grinnell College.

Innovative Ideas and Insights for Improving Program Resourcing across Seams

Edward Cardon, LTG (Ret.)—is an AIRC Research Fellow and Chair of the AIRC PPBE Integration Research Panel. He has over 36 years of military and executive management experience in military operations and establishing, leading, and transforming 14 organizations with diverse mission sets, including operations, education, cyber, and innovation. He transformed the Army Cyber Command and established other new cyber organizations. He was the Director of Business Transformation for the Army and led the task force that helped create the Army Futures Command. Since retirement, General Cardon serves on various advisory boards to improve national security. [edward.cardon@gmail.com]

Michael McGrath, DSc—is an AIRC Research Fellow and independent consultant with broad government and industry experience in technology, management, logistics, and acquisition. His prior senior DoD positions include DASN(RDT&E), ADUSD(Dual Use and Commercial Programs), a DARPA Program Manager, and OSD Director of the DoD Computer-Aided Acquisition and Logistics Support program. He has served on multiple Defense Science Board and National Academies studies. His research interests are in manufacturing, cybersecurity, and data science. McGrath holds a doctorate in operations research from George Washington University. [mmcgrat1@stevens.edu]

Hoong Yan See Tao, PhD—is a Research Project Manager in the Systems Engineering Research Center (SERC). Yan conducts and manages research in mission engineering, digital engineering, human capital development, systems engineering, and systems management transformation. She was on the Stevens Sustainability Committee and part of the student-led organizing committee for the inaugural Student Sustainability Symposium in 2016. In 2013, Yan was the President of the INCOSE Student Division at Stevens. Yan earned her BS and MS in electrical engineering from the University of Texas at El Paso and her PhD in systems engineering from the Stevens Institute of Technology. [hseetao@stevens.edu]

Abstract

The Department of Defense (DoD) has three decision support systems that must be synchronized to deliver capabilities at the right time: the Requirements; Acquisition; and Planning, Programming, Budgeting and Execution (PPBE) systems. PPBE is calendar driven, but both the requirements and acquisition processes, which are dependent on PPBE, are activity driven. Collectively, these three systems are sometimes called “Big A” Acquisition. This paper examines integration, interoperability, and interdependency issues at the seams among these systems. It summarizes research by a panel of experts convened in support of the PPBE Reform Commission. This research included over 50 discussion sessions with current and former executives from government, industry, and academia. This paper identifies key issues at the seams and offers recommendations to complement those of the PPBE Commission.

Introduction

Providing and managing financial resources is essential to our national security. However, the Planning, Programming, Budgeting, and Execution (PPBE) system falters in its ability to operate with the requisite velocity and flexibility to enable the Department of Defense (DoD) to keep pace with adversaries in the development and deployment of military capabilities. Recognizing this need for improvement, Congress chartered the PPBE Reform Commission (National Defense Authorization Act [NDAA], 2022) to develop recommendations for consideration by the DoD and Congress to improve PPBE. The Acquisition Innovation Research Center (AIRC) was, in turn, asked by the Commission to provide research inputs in several areas, including the integration of the DoD’s three major decision systems for delivering capabilities: the requirements, acquisition, and PPBE systems (see Cardon et al., 2023). While there have been numerous calls for change within these decision systems for decades, this paper focuses on the synchronization challenges across these systems. Delivery of capabilities to the warfighter hinges on integrating requirements development, resource allocation, and acquisition decisions. Consequently, enhancing synchronization among these systems is of paramount importance.

In this paper, we summarize issues that arise at three seams:

1. The PPBE-Acquisition Seam. This is the interface most directly addressed by the PPBE Reform Commission.
2. The Requirements-PPBE Seam. This interface is not well defined in current policies and practices.
3. The Requirements-Acquisition Seam. This interface is beyond the scope of the PPBE Reform Commission but is the target of other Congressional interest.

Methodology

Our methodology included a literature search of prior studies and analysis of issues identified by the AIRC Integration Research Panel, consisting of retired DoD general officers and senior executives: LTG (Ret.) Edward Cardon (chair), U.S. Army; David Drabkin, Esq. (co-chair); LTG (Ret.) Wendy Masiello, U.S. Air Force; LTG (Ret.) N. Ross Thompson III, U.S. Army; MG (Ret.) Robert M. “Bo” Dyess, U.S. Army; COL (Ret.) Michael Smith, U.S. Army; Elliott Branch; and Michael McGrath.

The panel met with 50 subject matter experts across the DoD, Services, industry, and academia on a not-for-attribution basis to garner insights and discuss better ways to synchronize across requirements, acquisition, and PPBE decision making processes to deliver better capability outcomes. Notes from these sessions were analyzed using a qualitative data analysis (QDA) process (QDA, 2024). Figure 1 summarizes the overall comments on PPBE, although the panel found specific comments to be of primary value to their findings and recommendations. Table 1 highlights some of the comments received in these discussions.

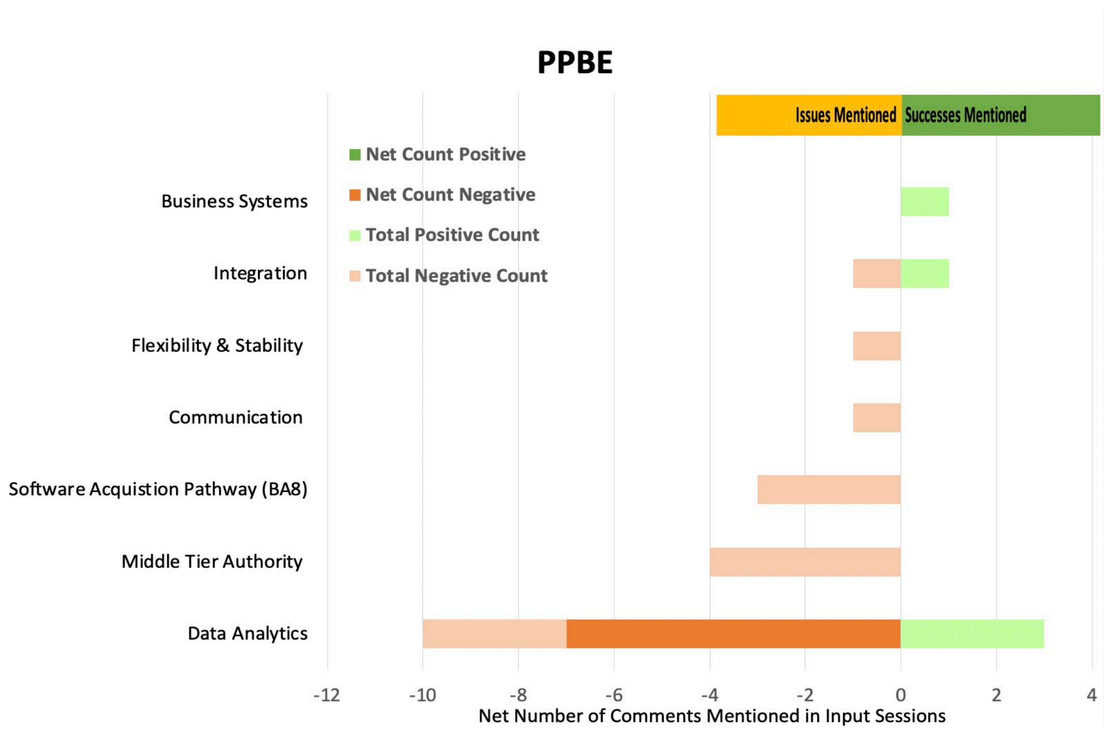


Figure 18. Categorization of Comments from the Input Sessions on the Overall PPBE

Table 1. Examples of Comments from Input Sessions

- PPBE is a good, rational, logical system – however, it is designed for a bi-polar world and not for the current environment.
- PPBE generally works well given the constraints.
- There is a lack of fidelity and granularity during initial planning that impacts Acquisition.
- We have a "Plan to Plan" but we don't have a "Plan to Decide."
- Lack of data availability and transparency hinders decision making.
- The Service programming process is overdesigned and unduly drives the process for strategy and acquisition.
- Program execution is a continuum, not a series of discrete budget executions.
- There is no streamlined approach to making changes once the budget is submitted.
- The reprogramming process is broken; it is almost impossible to get actions through four congressional committees in a timely fashion.
- The budget issue paper process leads to 3-star [general officer] meetings, where decisions are made with insufficient information.
- It's the volatility of budgets, not the performance of the industrial base, that is the [innovation] problem.
- Speaking of agility and flexibility, the DoD only has $4 billion in General Transfer Authority (GTA) out of an $855 billion budget [only about half of a percent]!
- It is difficult to plan and execute innovative research on a timeline.
- Too many involved are not accountable for results.

Panel Findings

The panel organized its conclusions and recommendations in three categories corresponding to the seams among the three decision support systems.

The PPBE/Acquisition Seam

PPBE is calendar-driven, but both the requirements and acquisition processes, which depend on PPBE, are activity driven. This disconnect is important because both requirements and acquisition have pathways and processes that have evolved to operate much more rapidly than the 2+ year long PPBE cycle. This temporal disconnect is especially critical for urgent needs and emerging technologies, wherein needs may arise more quickly. This disconnect causes delays and missed opportunities in the effort to develop and deliver timely warfighting capabilities to address rapidly evolving threats. A comprehensive analysis by the Hudson Institute (Greenwalt, 2021) concludes that time-based competition with our potential adversaries requires a holistic change in our resource allocation process.

The Joint Capabilities Integration and Development System (JCIDS) process generates validated capability requirements. Once a requirement is approved, resources are programmed in the PPBE process, and the execution shifts to the Defense Acquisition System (DAS), where Program Managers (PMs) are constrained by the annual cycle of when PPBE inputs can be made and when appropriations are subsequently issued. As acquisition events unfold, any delays in contracting, technology development, test and evaluation, and production problems may cause a mismatch between the acquisition plan and the availability of resources. This in turn may drive changes in resources by the DoD or Congress, but with a time lag that exacerbates the problem for the PM since many of the capabilities acquired need to be technologically current to meet the threat.

The current process is dependent on the calendars of appropriate senior leadership (multiple personnel at multiple levels for each decision), which adds further lags in the system. Additionally, the need to then coordinate with four congressional defense committees (House and Senate appropriations, House and Senate authorizations) for both initial inputs and adjustments in execution, using document-based inputs and interactions, is inefficient and ineffective in an era of rapidly changing technologies and threats.

The panel developed recommendations after reviewing the following munitions use-case example, technology transition problems, and issues of trust and transparency between PPBE and Acquisition.

Munitions Case Example

Munitions are often used as "bill payers" in the PPBE process. The U.S. Army is the DoD's Executive Agent for energetics, with an Organic Industrial Base (OIB)[1] that complements the commercial Defense Industrial Base (DIB). There are systemic issues with supply chain fragility, and there are current struggles to ramp up production of munitions due to operational needs and foreign aid. This situation has been obscured for years by faulty planning assumptions, optimization based on peacetime requirements, and complex chains of authority within the U.S. Army, the other Military Services, the Defense Agencies, and the Office of the Secretary of Defense (OSD). Recent responses to crises, such as the war in Ukraine and COVID-19, have revealed fragility in the ability to rapidly increase production. Over 50 mergers and acquisitions within the DIB have left five primes in control of the market, while inconsistent funding has discouraged industry investments. As a result, the DoD has seen a decline in production capacity over the past 30 years and lacks the surge capacity for several munitions it procures.

Defense-wide efforts on munitions procurement are affected by munitions requirements, budgeting, governance, and contracting. Formal processes are in place to establish munitions requirements, but senior leaders have little visibility into risks or tradeoffs. Munitions compete for modernization funds, which historically are then cut to pay for other bills based on an assumed ability of the DIB to surge capacity. The Army typically has no single authority that oversees end-to-end enterprise munitions matters, such as quantity and lethality requirements, the monitoring and mitigating of low demand signals to the OIB and DIB, the definition and establishment of minimum sustaining rates, the elimination of single points of failure, and the adjudication of disputes between munitions managers. Industrial concerns and constraints focus on the issues of sustainable procurement and capital investment. Industry partners uniformly complained about slowness of contracting and delayed investment decisions, while smaller businesses have been squeezed by inflationary concerns and uncertainty of future orders. Contracting personnel's compliance incentives do not align with their Ammunition Program Manager customers' mission focus, and the complexity of the Federal Acquisition Regulation (FAR)/Defense Federal Acquisition Regulation (DFAR) arrangements creates inefficiencies. Army-run facilities not only support Army munitions requirements and surge demands but also those of its sister Services.

Options for Improvements

There are several areas where the DoD can improve efforts in the near term. The U.S. Army could examine initiatives to strengthen unity of command, with the aim of simplifying control of munitions procurement and defining the roles of the PEOs and the Joint Munitions Command. The Army could define a single entity to establish requirements for new enhancements in lethality and range. The DoD could focus efforts on analyzing future strategic munitions needs to better prioritize availability for critical munitions with long lead times.

[1] "Organic" in that it resides within the government.

Addressing these issues could involve implementing larger (> $500 million) funding ceilings on multiyear procurement deals to establish minimum sustaining rates for munitions. Additionally, the Army should seek Congressional approval for purchasing long-lead items for critical munitions, facilitating future production surges. Moreover, expanding the use of more cost-effective and "attritable" munitions, suitable for training or foreign military sales, could be pursued. Lastly, the DoD could fund a flexible pilot plant line to explore methods of developing new explosive synthesis, jumpstart the adoption of new manufacturing technology, and ultimately create a model that would lessen reliance on foreign sources.

While these options and recommendations may mitigate much of the production risk exposed by demands stemming from the Ukraine conflict as we are aware of it today, the industrial bases (organic and commercial) may be incapable of meeting the munitions demand created by a potential future fight against a near-peer adversary. For example, a recent CSIS analysis of a hypothetical U.S. conflict with China (Jones, 2023) exposed significant shortfalls that go beyond what these recommendations could address.

Technology Transition -- the Valley of Death (VoD)

PPBE is sometimes blamed for technology-transition problems. The DoD invests heavily in technology innovation, but for a new technology to be transitioned into a program of record, the PM must have resources programmed and budgeted years in advance of transition. The panel found, however, that PPBE is always a matter of priorities. There are many examples of intervention by senior leaders and heroic efforts to reprogram funds to pull a technology across the valley of death (VoD), sometimes to meet an urgent need (like the MRAP program [Gansler et al., 2010]) and sometimes to provide a strategically important capability (like the Long-Range Anti-Ship Missile [LRASM; Defense Acquisition University (DAU), 2019]). There are also examples of programs that are structured in advance to include transition agreements and funding (such as the Future Naval Capabilities program). And there are examples of small 6.4 program elements that have budget-year flexibility to serve as bridge funding while the program of record arranges outyear funding. Ultimately, however, the PPBE process is intended to fund the highest priorities, and the argument is that any technology that lands in the VoD simply did not have the priority to make the cut. It should be noted that it takes an exceedingly high priority to instigate reprogramming that will "break" existing programs. If innovating to keep pace with potential adversaries is a priority, then maintaining BA 6.4 Program Elements (PEs) with flexible bridge funding would be less disruptive than reprogramming.[2]

Trust and Transparency

The panel's research suggests that an indispensable element to establishing transparency and trust within any complex system is direct, timely access to comprehensive and accurate data by the appropriate people and systems. There is a significant challenge in ensuring the transparency and accuracy of data within the PPBE processes due to issues that manifest in three distinct areas.

The first of these areas relates to the complex journey taken as data move both horizontally and vertically through various nodes of the decision-making hierarchies. Data originating from the Services, CCMDs, and various agencies must traverse a convoluted path as they progress from initial planning and programming stages to budgeting and execution phases. At each stage, the data are manually accessed, manually cross-referenced to operational service capability gaps (if not from a service), aggregated (generally in parts, rather than in whole), transformed (creating "new" data), and refined (creating more "new" data) to meet the specific data demands and formats of the respective entities involved—first within the

[2] This is consistent with the recommendation of the Defense Innovation Board to create "oasis" funding to bridge companies across the VoD. are often cited as the reason for technology projects getting stuck in the "valley of death" (VoD).

DoD, then with the Office of Management and Budget (OMB), and finally, with Congress. This intricate process of data transformation is often likened to an “information diode,” referring to a unidirectional flow of information, much like a one-way valve, and vividly illustrates the challenges at hand. By contrast, banking and investment firms use commercial technology and practices that incorporate necessary feedback loops, providing options for timely and well-informed decisions at a scale that are not possible today within PPBE’s processes.

The second formidable challenge lies in the pervasiveness of data silos, in which information is compartmentalized and opaque. The extensive nature of this challenge for PPBE is outlined in Section 7 of the PPBE Reform Commission’s (2024) Final Report. Our panel concluded that unrestricted access to all available data can profoundly enhance transparency and trust; however, it can also inadvertently lead to micromanagement and an unending deluge of inquiries concerning the purpose and outcomes of various activities. Navigating a balance requires that data access be judiciously granted to individuals and teams with direct responsibility and authority for making or informing critical decisions. By tailoring data access to align with specific roles and responsibilities, this middle ground would ensure that those entrusted with decision making possess the necessary information through coherent, real-time data visualization from a system (not from briefing charts or static documents) without being inundated with extraneous details.

The third area of concern is the allocation of decision rights and the establishment of clear accountability within the complex decision-making landscape. There is a compelling need for enhanced clarity in defining who holds the authority to make critical decisions and how those individuals are accountable for the outcomes of their decisions. As the Section 809 Panel (2019) discussed, teams at various levels should continue advising decision makers, but their operation must not delay the decision-making process with additional “sign offs.” If an issue remains unresolved, it should be elevated to the next decision level for resolution. There is a tendency to spend excessive time attempting to build consensus when the decision maker should simply consider all inputs, decide, and proceed. By comprehensively addressing the challenge of decision responsibility and accountability, PPBE organizations could bolster their capacity to make informed decisions, enhance operational efficiency, and cultivate a culture of transparency and trust with stakeholders.

A More Ideal Process for the PPBE-Acquisition Seam

The panel identified the following features needed in a more ideal PPBE process.

Planning – Could remain on an annual basis because it considers large groupings of resources in distant years. The inputs for this portion of the cycle are usually documents, such as the Defense Planning Guidance (DPG), National Defense Strategy (NDS), National Military Strategy (NMS), and the National Security Strategy (NSS), all of which look broadly and often at time horizons 3 to 30 years ahead. Critical within this phase is understanding the types of experimentation and testing that might be needed to fully provide novel technologies for the force so that resource demands can be identified.

Programming – Could remain on an annual basis. While its focus is not as broad as in planning, programming seeks to organize resources into logical groupings, and a higher confidence interval is placed on the expected needs. This phase requires the DoD and the Services to begin aligning resource needs to support anticipated high-level mission and portfolio demands roughly three years hence. However, it is unrealistic at this point to expect to know in detail (e.g., at the platform or system level) the solutions necessary to meet future capabilities. Thus, modernization programming should focus on groupings of capabilities that would capture aspects, such as technology development and maturation and operational experimentation, to understand the required capabilities more fully at organizational and platform levels. At the same time, programming must provide appropriate oversight by placing capabilities within context and with prioritization.

Budgeting – Could move to a semi-annual basis through a fixed, systematic mid-year review that provides a standard methodology for adjusting resources based on external factors and "fact of life" activities in emerging and established programs based on changes to threat, requirements, and technological breakthroughs. This phase has two discrete sub-elements, with the first portion focusing on the traditional assembly of budget documents that address individual program element level of detail, while the second focuses on realignment of resources based on fact-of-life adjustments.

Execution – Could move from calendar-based Comptroller sweeps of unobligated funds to event-based resource managers (S&T, R&D, or Acquisition) setting obligation schedules (plans) for each program when funds are appropriated, and DoD and Service Comptrollers measuring obligation status against these plans. Congress could maintain oversight through a data infrastructure that permits real-time monitoring of resources by Congress.

The Requirements/PPBE Seam

There is a major disconnect between the formal DoD requirements process and the PPBE process at every level below the Defense Planning Guidance (DPG). The Joint Requirements Oversight Council (JROC) validates joint capability development (DoD, 2021) but has little or no influence over PPBE priorities, which are set in the Service programming process. Inputs to the panel from Combatant Commanders (CCDRs) indicated a belief that their priorities are subordinate to Service priorities with no forum for resolution.

Industry needs more visibility into requirements to construct advanced manufacturing facilities, establish supply chains for long lead-time parts, or access advanced materials. This necessitates significantly earlier commitments from the DoD than currently exist within PPBE or acquisition contracts. Better government fidelity on requirements up to a threshold level (with options via spiral development to an objective level) covered by terminations clauses in contracts would reduce industry risk and by extension, risk-premium pricing.

The requirements process is the most under-resourced of the three major decision-making systems. While urgent requirements are approved quickly, the deliberate JCIDS process has been criticized as being too slow in practice, requiring 3–5 years to develop a validated requirement for a program of record (MITRE, 2020). The lack of fidelity on production numbers based on experimentation, simulation, and user touch points, combined with the lack of concepts, and use cases informed by the Services and CCMDs in conjunction with the S&T communities, creates unstable and unsettled requirements.

Organizational Focus

There has been growing tension with the delivery of capabilities among the CCMDs, Services, and Agencies. While the requirements process is ultimately intended to support the CCMDs, the Services are statutorily directed to develop capability to support the CCMDs in their specific domains. Hence, there have always been challenges in the development and integration of capabilities from the Services (and Agencies) to support the CCMDs. While one of the roles of the Joint Requirements Oversight Council (JROC) is joint capability development, the JROC itself does not have PPBE authorities.

The other tension within organizational design is the integration of commercial industry. The PPBE process is designed for a five-year plan/program supported with annual appropriations. By contrast, capital markets drive industry behavior with publicly traded companies focused on quarterly reports and annual forecasts. There are similar short-term pressures on companies supported through private equity firms, venture capital firms, or home offices. Given the pace of technologies, especially with the ever-increasing role of software in capabilities, industry lacks visibility of and confidence in the DoD requirements process. This problem is so acute that businesses factor this risk into price. This premium can be as much as 30%, directly impacting the top lines of the modernization portfolios and by extension the PPBE process. This is especially problematic for advanced manufacturing facilities and long lead time supplies of materials. The organization design of the PPBE processes is challenging (even antithetical) to commercial companies that operate in the dynamism of the capital markets.

Institutionally, a Cross-Functional Team (CFT) construct has proven highly successful in fostering integration on a large scale between the requirements and acquisition elements, providing enhanced efficiency and effectiveness within the current PPBE construct. While the DoD might find CFTs too unwieldy for widespread adoption across the Department, deploying CFT organizational structures strategically on the most critical (whether time or technology based) programs, featuring empowered leaders from all three decision-making systems and including appropriate Congressional representation to enable appropriate involved oversight, presents a promising solution to integration challenges. The USAF's B-21 program provides a prime exemplar of this approach, including integration with industry.

B-21 Use Case Example

The B-21 program, benefiting from its priority status within the Rapid Capabilities Office (RCO), enjoys significant advantages, including funding protection, priority resource allocation, and direct access to decision makers. This priority designation ensures that unobligated funds within the RCO portfolio, particularly those allocated to programs like the B-21, remain shielded from external budgetary pressures. The lean operational structure of the RCO, with a core team of fewer than 20 individuals, facilitates rapid decision-making processes, augmented as needed by user representatives such as Global Strike pilots, maintainers, and logisticians.

As a priority program, the B-21 receives attention from senior decision makers, allowing for timely discussions and issue resolutions. The program also benefits from direct access to key stakeholders, including Congress, the Office of Management and Budget (OMB), the Secretary of the Air Force (SECAF), and Senior Acquisition Officials (SAEs). Proximity to decision-making hubs in the DC area further enhances communication and fosters strong functional relationships, helping to address challenges effectively.

Trust plays a crucial role in the success of the B-21 program, with priority designation hinging on the establishment of trust through factual presentation, transparency, and good relationships. The RCO's approach of operating on facts, not opinions, and fostering transparency through first-hand knowledge and communication skills earns the confidence of decision makers and stakeholders. Moreover, building strong relationships with the user community, such as Global Strike Command representatives, enhances understanding of true options and needs.

The talent within the B-21 program is exceptional, with multifunctional teams selected through a rigorous selection process. User representatives, particularly from the Global Strike community, contribute significantly to shaping the program and bring valuable operational insights. The level of user support received by the B-21 program is remarkable, highlighting the program's commitment to excellence and collaboration.

In conclusion, the B-21 program's priority status within the RCO, coupled with its transparent and collaborative approach, fosters trust, enables efficient resource allocation, and ensures direct access to decision makers. The program's exceptional talent and strong user support further contribute to its success in meeting operational requirements and achieving mission objectives.

This B-21 user story offers ways to improve program performance from requirements determination to acquisition in partnership with the PPBE process. As a practical matter, not every program can have a top-priority designation nor have ready access to decision makers. Nevertheless, practices that would benefit acquisition include:

- Afford PEOs more funding flexibility among their portfolio programs while establishing accountability for success and transparency along the way.
- Establish smaller program teams to drive more firsthand involvement in program execution, thus increasing direct knowledge of progress and issues when engaging with stakeholders.
- Encourage multifunctional program offices to streamline acquisition processes and decisions.
- Co-locate user support with acquisition teams to accelerate requirements trades during the development and even production processes. User support might include operators, maintainers, logisticians, or other key non-traditional acquisition team members.
- Give CCMDs a greater voice in the requirements process. While Services have responsibility to organize, train, and equip in support of CCMDs, they still plan, program, budget, and acquire in stovepipes. CCMDs need a strong voice in today's interconnected realm of conflict.

The Requirements/Acquisition Seam

The panel recognized that this seam is outside the scope of the PPBE Reform Commission and therefore outside the panel's charter. Nonetheless, they found that it needs improvement and provide recommendations for future consideration. The current deliberate JCIDS process is widely criticized as too slow and bureaucratic to keep pace with technology or threat in cases where time matters. An AIRC (2022) report documented a sample of 20 Navy programs where the JCIDS staffing process took an average of 2.3 years to provide a validated Concept Development Document. JCIDS is based on an outmoded waterfall model rather than the highly iterative and collaborative agile development process now used in industry. Successful programs have used cross-functional teams for collaboration and iteration among requirement developers and system engineers, often with user representatives embedded in the program office (e.g., B-21).

The adaptive acquisition framework provides pathways for Middle Tier of Acquisition (MTA) and software development that are exempt from the JCIDS process and are being used successfully by DoD Components to develop and deliver new capabilities. The AIRC (2022) report found that these streamlined requirements validation processes have reduced the documentation and staffing times by 50% while still addressing joint needs. The FY 2024 National Defense Authorization Act (NDAA) Section 811 called for the DoD to modernize its requirements process using a "clean sheet" approach. In addition to acquisition reform and PPBE reform initiatives, reform of the requirements process is needed to achieve the agility the DoD needs.

Integration Across the Seams

Space Development Agency Use Case

The requirements, PPBE, and Acquisition processes have separate "process owners": VCJCS, USD(C),[3] and USD(A&S), respectively. Solving synchronization problems for acquisition purposes, therefore, typically falls to the PEO and PM with little help from the organizations above them. There are Service champions, such as the Deputy Assistant Secretary of the Navy (DASN) positions (e.g., DASN[Ships] or DASN[Air]) who facilitate the resolution of significant integration problems, but the routine integration tasks are managed at the program level. PEOs and PMs have become adept at using existing flexibilities and authorities to the maximum in resolving disconnects.

A good example is the Space Development Agency (SDA), a direct reporting unit of the U.S. Space Force. The SDA mission is to deliver needed space-based capabilities to the joint warfighter to support terrestrial missions through development, fielding, and operation of a proliferated low Earth orbit (pLEO) constellation of satellites. This Proliferated Warfighter Space Architecture (PWSA) program uses a spiral development strategy that is launching satellites in five tranches, with a new tranche every two years. Tranche 0 (FY 2022) satellites are successfully in orbit, and the program is on pace to deliver capabilities on schedule at a price point once deemed unachievable. This has been achieved through integration across the seams in the DoD decision systems:

- Acquisition uses the MTA pathway to go fast using Other Transaction Agreements (111 days from solicitation to contract award) for all but the ground segment of the system. An open architecture, a pool of qualified contractors, and competitive awards for each tranche keeps a warm base of innovation available. This process capitalizes on affordable, commercially available launch vehicles produced and launched by SpaceX. Spiral development allows adding capability as the threat evolves and provides flexibility to defer features to future tranches if they fall behind schedule or require additional investment. The limitations of MTAs, such as five years to fielding, fit comfortably within this program.
- PPBE provides funding in a single RDT&E appropriation that is used for both development and fielding. Changes in funding can be addressed by deferring or accelerating features in a tranche.
- MTA authority exempts the program from the JCIDS process, so requirements are approved by an SDA flag/SES Warfighter Council that meets semi-annually and is supported by monthly working groups. The Council includes representatives from the Services, CCMDs, S&T community, OSD, and other stakeholders. Requirements are directly reflected in the solicitations for each tranche.

This strategy has been highly effective to date in delivering initial capability to the warfighter. It is a delicate balance. Any changes in acquisition authorities, PPBE requirements to change to procurement funding, or assertion of JCIDS process compliance could reduce SDA flexibility and slow the pace of capability delivery to the joint warfighter. Nonetheless, the SDA has shown that integration of requirements, PPBE, and acquisition can be made to work. The changes recommended by the PPBE Reform Commission and our panel would make it easier for all programs to achieve comparable results.

[3] While each PPBE Phase has an owner, the USD(C)'s Program/Budget organization oversees the PPBE process.

Creating a Continuous Improvement Culture

The volume of information that has been compiled under the title of "Acquisition Reform" since the 1986 Goldwater-Nichols Department of Defense Reorganization Act is overwhelming. Notwithstanding the decades of documents, a back-to-basics approach is needed to continuously improve every aspect of the Department's Big A (requirements, resources, acquisition). In his book *Leading the Lean Enterprise Transformation*, George Koenigsaecker (2012) outlines what it takes to build a continuous improvement culture. The concept and practice of continuous improvement and the power of respect for people are the core principles. Every individual, in every organization, must be chartered with discovering the best way of doing everything, and every process employed in the DoD should be treated as purely experimental.

The result of continuously reviewing work is to define each step as either *value-adding work* or *non-value-adding work*. Value-adding steps *transform* something, either material in a production process or data in an administrative process. Non-value-adding steps, on the other hand, tend to move things around, involve rework, and do not contribute to warfighter capability outcomes.

Many of the initiatives currently underway have helped to align the DoD stakeholders around key areas of focus that will transcend the title of the initiative or the leader who championed it. The panel recommends elevating our perspective to look at the framework for how the DoD should continuously improve and recognize the capabilities required for high performing organizations. Four specific improvements offer high payoff for integrating across the "Big A" seams:

1. Requirements—Training the requirements community is a development precipitated by the 2007 NDAA. There are approximately two weeks of training offered by the DAU (one week online and one week in residence) that lead to certification for the requirements community. The panel recommended completing the coding of requirements billets across the DoD and then ensuring that the individuals filling those billets have the requisite training. This can be done for the Acquisition workforce through the DoD and Service DACMs.

2. Alignment of stakeholders (Requirements, Resources, and Acquisition) at every level for acquisition programs, not just at the most senior levels of the Department, is necessary to create the transparency required to ensure continuous process improvement and knowledge sharing. Transparency builds trust and fosters teamwork.

3. Responsibility, Accountability, and Authority—Individuals involved in the review or approval of a program should possess all three of these traits and capabilities to have a vote. There are many levels of review, and at every level there are people on the various staffs who do not add value toward transforming something in a material or administrative process. There is a short chain of command for PMs in the DoD—PM-PEO-SAE-DAE—that all have the requisite responsibility, accountability, and authority. This acquisition chain of command is the ideal way to leverage IPTs and CFTs, and that short chain of command should be duplicated for the requirements and resourcing communities. This reinforces the recommendation on stakeholder alignment.

4. Align on key metrics that are true enterprise-level metrics for each DoD process— Improvement targets should be >10% per year for each metric area, and improvements should be expected in four metric areas:

- Quality improvement
- Delivery/lead time/flow improvement
- Cost/productivity improvement
- Human development

Recommendations

The panel's input was cited in several places in the Commission's Final Report (PPBE Reform Commission, 2024). The authors of this paper agree that the final report's 28 recommendations, if implemented, will help considerably in resolving many of the issues addressed by our panel. In particular, the Commission recommendations in Table 2 will provide much needed flexibility and insight at the PPBE-Acquisition seam.

Table 2. PPBE Commission Recommendations Affecting the PPBE-Acquisition Seam

#5. Consolidate RDT&E Budget Activities (BA)
#6. Increase Availability of Operating Funds
#7. Modify Internal DoD Reprogramming Requirements
#8. Update Values for Below Threshold Reprogrammings (BTR)
#8A. Increase BTR Thresholds Based Upon the Nominal Growth of the Appropriation
#8B. Allow Reprogramming of a Small Percentage of an Entire Appropriations Account with Regular Congressional Briefings and Oversight
#8C. Simplify New Start Notifications by Increasing the Notification Threshold
#9. Mitigate problems caused by Continuing Resolutions (CR)
#10. Review and Consolidate Budget Line Items (BLI)
#11. Address Challenges with Colors of Money
#11A. Allow Procurement, RDT&E, or O&M to be used for the Full Cycle of Software Development, Acquisition, and Sustainment
#11B. Use O&M for Hardware Continuing Improvements
#11C. Align Program and Program Office Funding to the Predominant Activity of the Program
#12. Review and Update PPBE-Related Guidance Documents
#13. Improve Awareness of Technology Resourcing Authorities
#14. Establish Special Transfer Authority for Programs Around Milestone Decisions
#15. Rebaseline OSD Obligation and Expenditure Benchmarks
#16. Encourage Use of the Defense Modernization Account (DMA)
#17. Encourage Improved In-Person Communications
#18. Restructure the Justification Books (J-book)
#19. Establish Classified and Unclassified Communication Enclaves
#20. Create a Common Analytics Platform
#27A. Improve Training for Preparation of Budget Justification Materials
#27D. Improve Understanding of Private Sector Practices

Beyond the major recommendations of the Commission, there is an opportunity for the DoD to implement specific additional recommendations of our panel that went beyond the Commission's scope. We summarize these additional recommendations by the synchronization seam they affect.

PPBE/Acquisition Seam Recommendations

- To reduce the time for integration from a PPBE perspective, the DoD should define clear roles and responsibility (who can say "yes," and more importantly, limiting who can say "no" to approvals) and avoid the drive for consensus through staff action by elevating issues to decision makers in a timely manner; For example, on the acquisition side, it is recognized that the top line for every program is a prioritization function that comes out of a larger PPBE process. Once that top line decision is made, the policy should clearly state that:

- only the PEO has approval authority over the PM from program perspectives; all others are advisory to the PM and PEO but cannot nonconcur.
- only the Component Acquisition Executive (CAE) has approval authority over the PEO; all others are advisory to the PM and PEO but cannot nonconcur.
- only the Defense Acquisition Executive (DAE) has approval authority over the CAE; all others are advisory to the PM and PEO but cannot nonconcur.
- the Milestone Decision Authority (MDA) is the main stopping point for approvals up the acquisition chain-of-command; the policy clearly states that "For MDAPs, it is DoD policy to budget to the DCAPE ICE unless an alternative estimate is specifically approved by the MDA"—thus, no others have an ability to say "no"; and
- those above the MDA in the acquisition chain-of-command can intervene in oversight, but this should be minimized.

- The DoD should link the concept of affordability in PPBE (DoDD 7045.14, Enclosure 3) to the affordability analysis called for and defined in the acquisition community (DoDI 5000.85, Section 3, and underlying processes). Affordability analysis results should be provided to inform all JCIDS requirements validations.

Requirements/PPBE Seam Recommendations

- The DoD should empower the JROC to assign a validated CCMD Joint Emerging Operational Need Statement (JEONS) to a Service or Agency as a "must fund" priority, with DEPSECDEF visibility of the resulting resource decisions. Require that CCMDs prioritize their requirements as part of the JROC requirements validation process, and that requirement lists be matched to and reconciled with Service Budget requests in the PPBE process by DEPSECDEF.
- The Joint Staff and the DoD should give CCDR-provided scenarios, exercise, and wargaming results weight equal to that given to the Military Services and Joint Staff inputs as the basis for the annual Capability Gap Analysis of the Future Years Defense Program (FYDP).
- The DoD should provide Service affordability analysis along with requirements that are reviewed and approved by the JROC. This will provide the JROC with the Service's sense of priorities and affordability with respect to the materiel item in question. Affordability analysis is required at Milestone A and thus is available for CDD validation (see DoDI 5000.85).
- To provide Industry more visibility into DoD requirements, especially with respect to production capacity, the DoD should include in budget justification documents provided publicly with the President's budget request both a threshold [minimum] and an objective [stretch goal] level for annual procurement quantities. DoD acquisition programs should reflect these requirements with contract options to the objective level and termination liability clauses applicable below the threshold level. In addition, the DoD should provide cleared defense contractors with controlled access to validated mission needs and requirements statements (at the CUI and classified levels) to help with industry's planning for Internal Research and Development (IR&D), staffing, and infrastructure investments and investment hedges.
- The DoD should provide cleared Industry (along with Congress) data and information from the President's Budget justification books in structured machine-readable formats. (This will also facilitate improved data analytics and portfolio views discussed in other AIRC reports to the PPBE Commission.)

Requirements/Acquisition Seam Recommendations

- The panel agreed with the FY 2024 NDAA Section 811 direction to reform the DoD requirements system. It recommended starting now on such reforms, to include:
    - Forming a JS-led CFT with OSD and Service stakeholders to reform the system, specifically the boundary between Requirements (JCIDS) and Acquisition (Defense Acquisition System [DAS]).
    - Developing a more agile, collaborative, and iterative process for the integration and transition of requirements to the systems engineering process.
    - Developing a capability needs and requirements framework with pathways that are aligned to the Department's Adaptive Acquisition Framework. This would include insight into the Department's S&T processes to identify emerging products that address capability requirements.
    - Developing a process to rapidly validate the military utility of commercial solutions to meet capability needs or opportunities.
    - Developing a mission engineering approach for defining enduring requirements in a set of capability portfolios, with a set of mission impact measures that capability deliveries must seek to continuously improve.
    - Assessing best practices to ensure that the requirements process for software, artificial intelligence, data, and related capability areas enable a more rapid, dynamic, and iterative approach than used for hardware systems.
    - Developing a formal career path, structure, and training for professional requirements managers.
- In addition, the panel recommended that the reforms of the DoD Requirements process include designating a single organization or entity directly responsible for overseeing and driving the development of joint capabilities identified as CCMD priorities.

Topics for Further Research

The panel identified several promising ideas and potential recommendations that require more research or prototyping before they can be finalized.

- Existing technology can be used for a rapid prototype of a Large Language Model (LLM)-enabled approach to J-books. Commercial offerings allow the DoD to select whatever LLM is best suited (and replace it when something better is available), use controlled DoD data sources for training the model, guarantee factual accuracy and citable sources without risk of hallucinations, and demonstrate the utility of the system in responding to complex natural language queries. We believe a spiral prototype interacting with users can validate key aspects of the system well within a year.
- Budget execution reviews could move from calendar-based Comptroller sweeps of unobligated funds to acquisition managers setting an event-based obligation schedule for each program when funds are appropriated, and DoD and Service Comptrollers measuring obligation status against these schedules. Congress could maintain oversight through a data management infrastructure that permits near real-time monitoring of execution status. Needed research includes further investigation of historical obligation patterns on acquisition programs compared to the normal linear execution model.
- The DoD could ask the geographic CCMDs to propose regional equivalents to the European Deterrence Initiative (a good example) for consideration in future planning and

programming. The CCMDs and associated Service funding lines would have to prioritize within available dollars and then engage in the program and budget review processes for additional resources, if required. The CCMDs should use the capability in the Services/Agencies to execute the funds for the CCMD priorities rather than duplicate program offices, contracting, etc. That gives the CCMDs more flexibility than waiting to the end of the POM to see how their IPLs stacked up for funding. It also incentivizes the Services for meeting CCMD IPL requirements with increased funding. If a more radical approach is possible, geographic CCMDs might be given substantial control over funds for Joint emerging needs. Research is needed to develop a method of cross-CCMD coordination to avoid duplication of capability development efforts, to get stakeholder views, and to provide cost estimates. A CFT with CCMD, Service, OSD, and JS representation would be needed.

- To better inform industry on production capacity planning, the DoD could provide access to Defense Contract Management Agency (DCMA) and Defense Logistics Agency (DLA) supply chain insights to better recognize, plan, and fund for supply chain risks and production capacity issues on highest priority, cross-program parts, and end-of-life procurement needs. This would need further research regarding protection of proprietary interests and analysis of the differences between production and sustainment supply chains.

Conclusion

The AIRC Integration Research Panel is deeply honored to have supported the PPBE Commission and its recommendations. This document encapsulates our support for and endorsement of the PPBE Commission's efforts. The recommendations outlined in the PPBE Report will enhance the decision-making processes critical to delivering capabilities to the DoD. In addition, while some recommendations were not accepted, and others exceeded the mandate of the PPBE Commission, further analytical scrutiny of these recommendations by the Acquisition Research community could yield significant enhancements to the "Big A" decision-making framework essential for delivering capabilities to the DoD.

References

AIRC. (2022). *Joint Capabilities Integration and Development System (JCIDS).* https://acqirc.org/publications/research/joint-capabilities-integration-and-development-system-jcids/

ASB. (2023). *Surge capacity in the defense munitions industrial base.* https://asb.army.mil/Portals/105/Reports/2020s/ASB%20FY%2023%20DMIB%20Report%20(E).pdf

Cardon, E., Drabkin, D., Branch, E., Dyess, R. M., Masiello, W., Smith, M. … Anton, P. S. (2023, December 22). *Improving defense outcomes through improved interfaces between PPBE, acquisition, and requirements* (AIRC-2023-TR-023). Acquisition Innovation Research Center. https://acqirc.org/publications

Defense Acquisition University. (2019). *Full speed to the fleet*. https://www.dau.edu/datl/b/full-speed-fleet

Defense Innovation Board. (2023). *Terraforming the valley of death*. https://innovation.defense.gov/Portals/63/DIB_Terraforming%20the%20Valley%20of%20Death_230717_1.pdf

DoD. (2021). *Charter of the Joint Requirements Oversight Council and implementation of the Joint Capabilities Integration and Development System* (CJCSI 5123.01). https://www.jcs.mil/Portals/36/Documents/Library/Instructions/CJCSI%205123.01I.pdf

Gansler, J., et al. (2010). *Acquisition of mine-resistant, ambush-protected (MRAP) vehicles: A case study*. https://calhoun.nps.edu/handle/10945/33544

Greenwalt, W. (2021). *Competing in time*. Hudson Institute. https://www.aei.org/wp-content/uploads/2021/02/Greenwalt_Competing-in-Time.pdf

Jones, S. (2023). *Empty bins in a wartime environment: The challenge to the U.S. defense industrial base.* CSIS. https://csis-website-prod.s3.amazonaws.com/s3fs-public/2023-01/230119_Jones_Empty_Bins.pdf

Koenigsaecker, G. (2012). *Leading the Lean enterprise transformation* (2nd ed.). McGraw-Hill.

MITRE. (2020). https://www.mitre.org/sites/default/files/2021-11/prs-19-03715-2-modernizing-dod-requirements-enabling-speed-agility-and-innovation.pdf

National Defense Authorization Act for Fiscal Year 2022, Pub. L. No. 117-81, § 1004 (2022). https://www.congress.gov/bill/117th-congress/senate-bill/1605.

PPBE Reform Commission. (2023, August). *Interim report*. https://ppbereform.senate.gov/interimreport/

PPBE Reform Commission. (2024, March). *Final report*. https://ppbereform.senate.gov/finalreport/

QDA. (2024). *Qualitative research*. https://en.wikipedia.org/wiki/Qualitative_research

Section 809 Panel. (2019, January). *Streamlining and codifying acquisition regulations* (Vol. 3).

Research and Development: DOD Benefited from Financial Flexibilities but Could Do More to Maximize Their Use

Leslie Ashton—is a Senior Analyst at U.S. Government Accountability Office (GAO) [ashtonl@gao.gov]

Abstract

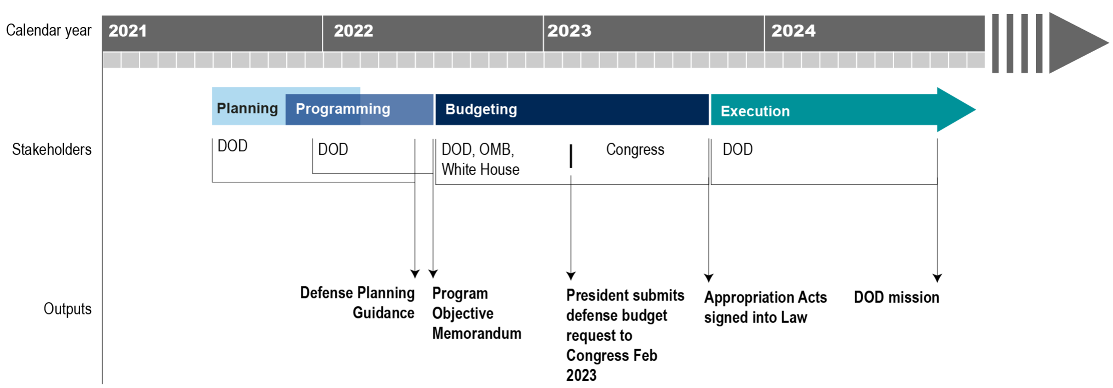
DOD receives about $95 billion annually to fund its R&D activities. Members of Congress, DOD officials, industry representatives, and researchers have raised questions about whether the process used to request and allocate those funds is fast and flexible enough to respond to evolving threats. GAO examined authorities related to budgeting and financial management that allowed DOD flexibility in its use of funds to support R&D, innovation, and modernization activities during fiscal years 2017 through 2021. The presentation will provide an overview of the 26 flexibilities GAO identified and will focus on DOD's use of five to accelerate R&D efforts. These five flexibilities supported thousands of activities contributing to DOD R&D and efforts to modernize or innovate capabilities for military departments. GAO found three factors—planning, guidance, and institutional support—that helped enable DOD officials' use of the five flexibilities. GAO reviewed U.S. Code, relevant legislation, and DOD documents; selected a nongeneralizable sample of five flexibilities, chosen to provide variation in what they allowed DOD to do, and 25 activities as illustrative examples and to assess their use; and interviewed DOD and military department officials.

Keywords: Research and development, Modernization, Innovation, Acquisition authorities, Planning Programming Budgeting and Execution (PPBE) process

Background

DOD decides how much funding to request annually for each military department through the PPBE process. According to DOD, the objective of the process is to provide the department with the most effective mix of forces, equipment, personnel, and support attainable within fiscal constraints. It involves numerous offices within DOD and the military departments; the Office of Management and Budget; the White House; and Congress.

The process begins with strategic planning and ends with the execution, or obligation, and expenditure of funds to complete DOD's mission, such as developing and delivering technologies to the warfighter. It generally takes around 2 years to obtain funding but can take longer (see fig. 1).



Source: GAO analysis of Department of Defense (DOD), GAO, and Office of Management and Budget (OMB) information. | GAO-23-105822

Figure 1. Notional DOD Planning, Programming, Budgeting, and Execution Timeline, Phases, Stakeholders, and Outputs for Fiscal Year 2024 Funding

In June 2017, we reported that the lengthy PPBE process can slow innovation. For example, a project conceived in November 2021 might not be authorized and appropriated funding until October 2023 or later. Projects that are expected to take 3 to 5 years to complete in effect can require 5 to 7 years from conception to completion. We also reported that these long timelines can make it difficult to achieve the adaptability and faster capability development and fielding times that DOD seeks to keep pace with rapidly evolving threats.

Additionally, over the last 10 years, budget submissions and appropriations acts have generally been late. On average budget submissions were 42 days late and appropriations acts were signed into law 108 days after the fiscal year start. Leaders from both the executive and legislative branches have identified lengthy delays in regular appropriations as a threat to national security. In addition, some of these leaders have publicly stated that the delays contribute to ineffective use of funds.

Annual Defense Appropriations

During the budget phase of the PPBE process, Congress drafts legislation that, when signed into law, provides DOD with budget authority in appropriations acts. Congress specifies the purpose for which each appropriation may be used, the amount of budget authority available, and the time period in which it is available under each appropriation. DOD uses that authority during the final phase of the PPBE process to execute its mission. Most of DOD's appropriations can be grouped into five major categories. Appropriations may be used only for their intended purposes and, for fixed-period appropriations, only for a defined period of time. See table 1 for examples of the four categories of appropriations included in this report; the fifth is for Military Personnel. Two of the appropriations categories—RDT&E and Military Construction (MILCON)—are discussed further below.

Table 1: Selected Categories of Defense Appropriations

Appropriation category	Notional examples of use	Years available for new obligation
Research, Development, Test and Evaluation	Funds activities performed by government laboratories, universities and contractors for the research and development of equipment and software, and its test and evaluation	2
Procurement	Funds acquisition programs approved for production and the costs integral to delivering a useful end item intended for operational use or inventory, including purchase of software licenses	3
Operation and Maintenance	Funds civilian salaries, travel, software license renewals, minor construction projects, training and education, depot maintenance, operating military forces, and base operations support	1
Military Construction	Funds major construction projects such as bases, schools, missile storage facilities, medical/dental clinics, military family housing, sensitive compartmented information facilities, and research and development installations	5

Source: GAO summary of Department of Defense information (Financial Management Regulation, 7000.14-R). | GAO-23-105822

To maintain technological superiority on the battlefield, DOD relies on scientific and technical knowledge developed largely through R&D activities and investments funded by the department and performed by industry, universities, government labs, and others. RDT&E appropriations include eight budget activities and largely fund DOD's R&D efforts. For example, the first three budget activities generally represent efforts undertaken by research laboratories, industry, and academia to advance research in areas important to U.S. military capabilities, drive long-term innovation, and develop technology. The other five budget activities are typically associated with product development for acquisition programs or fielded capabilities and comprise the majority of RDT&E funds.

Many organizations within DOD are involved in R&D activities, from setting priorities to execution and oversight. See figure 2 for examples of the stakeholders involved in science and technology funding.

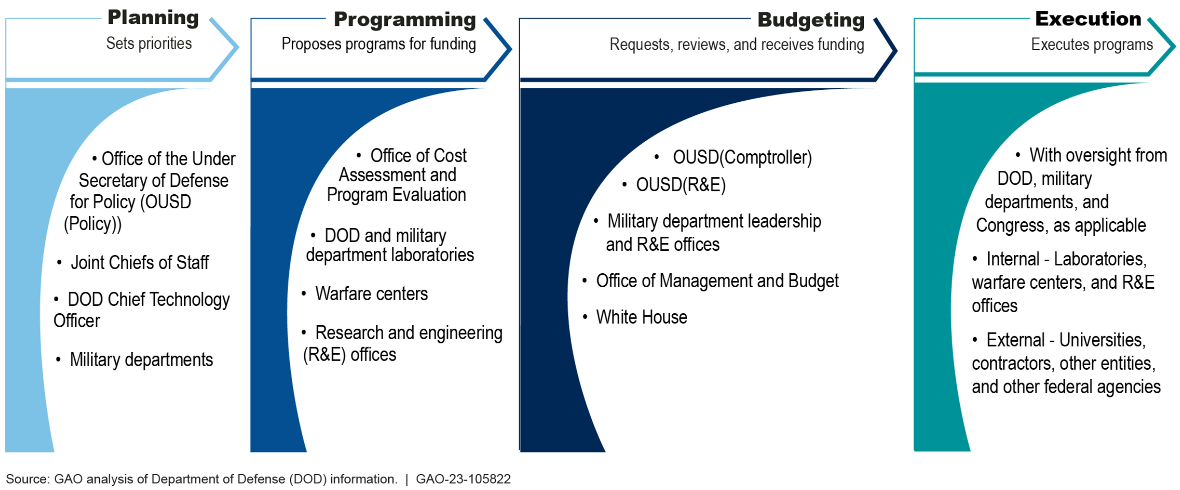


Figure 2. Examples of DOD Science and Technology Stakeholders in the Planning, Programming, Budgeting, and Execution (PPBE) Process

Some of the key officials and organizations involved in the implementation and oversight of R&D-related efforts include:

- The Under Secretary of Defense for Research and Engineering (USD(R&E))—the principal advisor to the Secretary of Defense for research, engineering, and technology development activities and programs—serves as DOD's chief technology officer. The powers and duties of this office include establishing policies and providing oversight for DOD's research, engineering, and technology development activities.
- The Deputy Chief Technology Officer for Science and Technology supports DOD's research and engineering mission by helping to ensure comprehensive, department-level insight into the activities and capabilities of the defense labs. The Deputy Chief's office carries out a range of core functions related to the defense labs, including analysis of capabilities, alignment of activities, and advocacy.
- The USD(Comptroller)—the principal advisor to the Secretary of Defense for budgetary and fiscal matters—serves as DOD's chief financial officer and administers the budget and execution phases of the PPBE process. The powers and duties of this office include financial management, accounting policy and systems, budget formulation and execution, and contract and audit administration.
- The USD for Acquisition and Sustainment (A&S)—the principal advisor to the Secretary of Defense for all matters relating to acquisition and sustainment, including system design and development; production; installation maintenance, management, and resilience; military construction; and procurement of goods and services, among other things. The powers and duties of this office include establishing policies and providing oversight of the DOD acquisition system, including rapid acquisition policies for urgent operational needs and acquisition of software.
- Military Department Assistant Secretaries of Air Force, Army, and Navy responsible for acquisition, technology, and logistics generally oversee, or have responsibilities related to, R&D. The powers and duties of these offices generally include establishing policies and providing oversight for the military departments' research, engineering, technology development, and acquisition activities.

- Military Department Assistant Secretaries of the Air Force, Army, and Navy responsible for financial management serve as comptrollers of the military departments. They are responsible for policies, procedures, programs, and systems pertaining to finance and accounting activities and operation. The powers and duties of these offices generally include RDT&E budget formulation, the presentation and defense of the budget through the congressional appropriation process, budget execution and analysis, reprogramming actions, and appropriation fund control/distribution.
- Military Department Laboratories conduct R&D activities along with universities, federally-funded research and development centers, and other entities.

Military Construction

MILCON funds R&D-related construction projects, including facility modernization and new construction, among other things. R&D-related construction projects represent a relatively small proportion of needs and compete for funding with other construction projects, such as runways, piers, barracks, schools, hospitals, and other facilities. DOD includes a fraction of its construction-related needs each year in the President's budget request, which can result in neglected facilities that become more costly to maintain and repair. For example, the maintenance portion of the fiscal year 2022 budget request for MILCON included $348 million whereas a couple years earlier, for fiscal year 2020, DOD reported a deferred maintenance backlog of $137 billion. DOD leadership has raised concerns about the performance, reliability, and long-term viability of DOD's lab and test center infrastructure given the degraded facilities. To assist DOD labs, Congress has authorized certain flexibilities to help address laboratory construction and maintenance needs.

Flexibilities

Congress generally provides defense budget oversight, direction, and authorities to DOD through two annual bills—defense appropriations and authorization acts. Once signed into law, some of these legislative authorities allow DOD to address problems the department or Congress identified by providing DOD with financial flexibility in use of funds to support R&D, innovation, and modernization activities. These financial flexibilities may be limited to relatively small amounts of funding or target high-priority activities, such as addressing improvised explosive devices. Congress can provide temporary financial flexibilities, such as a pilot program for a new budget activity during which DOD and Congress can learn how a change in operations may work without investing relatively large amounts of funds or committing to long-term changes.

Congress can also give DOD the discretion to exercise a financial flexibility or not. For example, Congress authorized the Pilot Program on Modernization and Fielding of Electromagnetic Spectrum Warfare Systems and Electronic Warfare Capabilities in fiscal year 2017, and it remains in effect through fiscal year 2023. However, in May 2018, DOD notified the House of Representatives Committee on Armed Services that it had not established the pilot program because it would instead use modernization plans to improve legacy electromagnetic spectrum warfare and electronic warfare systems.

Financial flexibilities can also vary in terms of the discretion granted to DOD as demonstrated by the five flexibilities examined further in this review.

1. **Funding Laboratory Enhancements Across Four Categories (FLEX-4).** First introduced in fiscal year 2009 and codified in legislation enacted in 2017, this flexibility requires DOD to establish mechanisms for labs to use certain funds. In the event the director of a lab decides to use the flexibility, they must use between two and four percent

of the lab's available funds for basic and applied research, workforce development, efforts that support transitioning technology into operational use, and lab repair, revitalization, and minor refurbishment activities. The military departments have internal procedures for determining how to spend these funds but do not have to go through the full PPBE process.

2. **Defense Research and Development Rapid Innovation Program, also known as Defense Rapid Innovation Fund (RIF)**. First introduced in fiscal year 2011 and codified in legislation enacted in 2018, RIF allows DOD to transfer RIF-available funds to department RDT&E appropriations accounts (e.g. from Defense-wide RDT&E to Army RDT&E) to develop innovative technologies. RIF activities focus on maturing and demonstrating technologies in a relevant environment with the goal of transitioning them to defense programs.
3. **Rapid Acquisition Authority (RAA)**. First introduced in fiscal year 2003 and codified in legislation enacted in 2022, RAA allows DOD to use any of its available funds for the urgent acquisition and deployment of capabilities to eliminate deficiencies that could result in mission failure or loss of life. The funding decisions are approved within the department and do not have to go through the planning, programming, and budgeting phases of the PPBE process, but DOD must notify Congress about its use.
4. **Software and Digital Technology Pilot Programs, also known as Budget Activity Eight (BA-8)**. Introduced in fiscal year 2021, this pilot, using a new RDT&E budget activity, allows certain DOD programs to develop, buy, and maintain software using a single appropriation category rather than the three appropriations categories typically required for these types of efforts (RDT&E, Procurement, and Operation and Maintenance [O&M]).
5. **Defense Laboratory Modernization Program (Lab Modernization)**. First introduced in fiscal year 2016 and codified by the James M. Inhofe National Defense Authorization Act for Fiscal Year 2023, Lab Modernization allows DOD to obligate RDT&E, rather than MILCON, funds to support certain lab- or test center-related military construction. DOD must comply with military construction and congressional notification requirements.

DOD Has Not Communicated Information about Available Financial Flexibilities across the Department

We found that DOD has not broadly communicated information about available financial flexibilities throughout the agency. The Office of the Under Secretary for Defense (OUSD)(R&E), OUSD(Comptroller), and officials from the military departments responsible for research do not maintain centralized information on financial flexibilities that can be used to support DOD's R&D, innovation, and modernization efforts, nor is there a single responsible organization for these flexibilities. Instead, responsibility is distributed across different organizations in the department. OUSD(R&E) and OUSD(Comptroller) officials said that makes compiling information on the flexibilities difficult.

Without centralized information on financial flexibilities, we took steps to identify financial flexibilities available to DOD during fiscal years 2017 to 2021 to support its R&D efforts. The 26 financial flexibilities we identified support: (1) laboratory and test facility needs; (2) technology development; (3) development and fielding of capabilities to address specific threats; or (4) modern software development. We found over half of the 26 flexibilities provided DOD with decision-making over funds that it collected from providing services or that nonfederal government entities contributed towards certain DOD project costs. Table 2 lists the 26 financial flexibilities we identified. Appendix I provides summaries, congressional reporting requirements, and other information about these flexibilities. There may be additional flexibilities that are not included, but this resource may be a helpful starting point.

Table 2: Financial Flexibilities Relevant to DOD's Research and Development, Innovation, and Modernization Efforts from Fiscal Years 2017 through 2021

Flexibility	United States Code (U.S.C.) or legislation	Fiscal year originally authorized
Supports laboratory and test facility needs		
Availability of Samples, Drawings, Information, Equipment, Materials, and Certain Services	10 U.S.C. § 4892	1994
Centers for Science, Technology, and Engineering Partnership	10 U.S.C. § 4124	2016
Cooperative Agreements for Reciprocal Use of Test Facilities: Foreign Countries and International Organizations	10 U.S.C. § 2350l	2002
Defense Laboratory Modernization Program[a]	10 U.S.C. § 2805(g)	2016
Enhanced Transfer of Technology Developed at Department of Defense (DOD) Laboratories	National Defense Authorization Act (NDAA) for Fiscal Year (FY) 2014, Pub. L. No. 113-66, § 801 (2013), as amended, (10 U.S.C. § 4832 note)	2014
Federal Defense Laboratory Diversification Program	10 U.S.C. § 4833	1995
Mechanism to Provide Funds for Defense Laboratories for Research and Development of Technologies for Military Missions. DOD refers to this flexibility as the Funding Laboratory Enhancements Across Four Categories (FLEX-4)a	10 U.S.C. § 4123	2009
Pilot Program to Improve Incentives for Technology Transfer from DOD Laboratories	NDAA for FY 2018, Pub. L. No. 115-91, § 233 (2017), as amended (10 U.S.C. § 4832 note)	2018
Unspecified Minor Construction – Laboratory Revitalization	10 U.S.C. § 2805(d)	1982
Use of Test and Evaluation Installations by Commercial Entities	10 U.S.C. § 4175	1994
Supports technology development		
Authority of the Department of Defense to Carry Out Certain Prototype Projects	10 U.S.C. § 4022	2016
Cooperative Research and Development Agreements: North Atlantic Treaty Organization (NATO) Organizations; Allied and Friendly Foreign Countries	10 U.S.C. § 2350a	1990
Defense Dual-use Critical Technology Program	10 U.S.C. § 4831	1993
Defense Research and Development Rapid Innovation Program. DOD refers to this flexibility as the Defense Rapid Innovation Fund (RIF)a	10 U.S.C. § 4061	2011
Foreign Contributions for Cooperative Projects	10 U.S.C. § 2350i	1992
Manufacturing Technology Program	10 U.S.C. § 4841	1994
Military Aviation and Installation Assurance Clearinghouse for Review of Mission Obstructions	10 U.S.C. § 183a	2018
Nontraditional and Small Contractor Innovation Prototyping Programb	NDAA for FY 2017, Pub. L. No. 114-328, § 884 (2016), as amended	2017
Prizes for Advanced Technology Achievements	10 U.S.C. § 4025	2000
Rapid Prototyping Fundc	NDAA for FY 2016, Pub. L. No. 114-92, § 804(d) (2015), as amended	2016
Research Projects: Transactions Other than Contracts and Grants	10 U.S.C. § 4021	1990

Supports development and fielding of capabilities to address specific threats		
Joint Improvised-Threat Defeat Fund	Consolidated Appropriations Act, 2017, Pub. L. No. 115-31, Div. C, Title IX, Other Department of Defense Programs (2017)	2017
National Defense Sealift Fund	10 U.S.C. § 2218	1993
Pilot Program on Modernization and Fielding of Electromagnetic Spectrum Warfare Systems and Electronic Warfare Capabilities	NDAA for FY 2017, Pub. L. No. 114-328, § 234 (2016) (10 U.S.C. § 113 note)	2017
Procedures for Urgent Acquisition and Deployment of Capabilities Needed in Response to Urgent Operational Needs or Vital National Security Interest. DOD refers to this flexibility as the Rapid Acquisition Authority (RAA)	10 U.S.C. § 3601	2003
Supports modern software development		
Software and Digital Technology Pilot Programs. DOD refers to this flexibility as Budget Activity Eight (BA-8)	Consolidated Appropriation Acts, 2021, Pub. L. No. 116-260, § 8131 (2020)	2021

We found that some Army, Navy, and Air Force officials who are responsible for, or work at, department-level R&D organizations were not familiar with certain flexibilities for technology development and technology transfer. OUSD(R&E), OUSD(Comptroller), and some military department officials explained that an official might not be familiar with some flexibilities because they might be new to their roles, the flexibility is not widely used, or the flexibility does not pertain to their area of responsibility. Some of these officials explained that while a senior official responsible for R&D efforts or laboratory director might not be aware of all available financial flexibilities, they could rely on their staff to provide information about various flexibilities and advocate for use of the flexibility to meet a research need. However, some of these officials stated that leadership's lack of familiarity with financial flexibilities could lead to underuse of flexibilities to support DOD's R&D efforts.

- An Army official responsible for laboratory management and Navy officials responsible for R&D policy said that they review legislation, such as National Defense Authorization Acts and appropriation acts, to identify relevant flexibilities to their area of responsibility, and pass this information to officials within their chain of command.
- Officials in the Office of the Secretary of the Air Force said that they reviewed guidance issued by the Office of Management and Budget and DOD to identify relevant flexibilities. Furthermore, they explained, many program element monitors—officials responsible for a specific program in the budget request—annually review the National Defense Authorization Act to identify applicable flexibilities for their program. Officials said the review is a time-consuming task on top of their primary responsibilities.
- A senior OUSD(R&E) official stated that they collect legislative information about the flexibilities under their purview and share information with lab governance panels. However, we found the information was generally related to hiring authorities rather than financial flexibilities.
- OUSD(Comptroller) publishes a summary of certain flexibilities on its website. However, it was not comprehensive and included two of the 26 flexibilities we identified. In addition, some officials responsible for R&D and financial management said they were not familiar with this resource. OUSD(Comptroller) officials explained this resource generally covers flexibilities that involve their office, such as reprogramming and transfer authorities. The summary is from January 2021, but a responsible official said that they plan to update it in 2023.

However, these officials stated that this information is not necessarily widely available. For example, information OUSD(R&E) collects and shares would not be available to officials who are not part of OUSD(R&E)'s governance panels. As of January 2023, an OUSD(R&E) official stated that they are considering whether to make the information on flexibilities they track available department-wide but do not have a specific timeline for when this would be completed.

In addition, the annual reviews of new legislation may not result in a complete understanding of the breadth of available flexibilities. The reviews do not capture financial flexibilities that are not amended annually. For example, an official could miss the availability of the Enhanced Transfer of Technology Developed at DOD Laboratories flexibility because it was infrequently amended—once in 2016 and 5 years later in 2021. As of March 2023, the flexibility has not been amended.

Army, Navy, and Air Force officials said that having widely available information about the financial flexibilities would be helpful to confirm their understanding of the flexibilities and to ensure they did not miss identifying relevant authorities that could support R&D efforts. For example, they said that a resource with information about a flexibility— such as whether it identifies a funding source, is authorized for a fixed period of time, and has congressional reporting requirements—would help them understand how to use it. In addition, identifying relevant DOD and military department guidance would help facilitate the use of financial flexibilities, according to Army and Navy officials. Air Force officials said that DOD could use existing mechanisms to widely communicate information about the flexibilities, such as having Defense Acquisition University courses cover current financial flexibilities or refer to a resource with such information.

Standards for Internal Control in the Federal Government calls for management to internally communicate the necessary quality information to achieve an objective. Similarly, we previously reported that DOD should educate users to maximize the use of flexibilities to address various challenges, ranging from quickly fielding solutions to the warfighter to increasing innovation from nontraditional defense contractors. We also reported that DOD increased its use of human capital authorities as the agency's leadership encouraged the use of the authorities and provided guidance to address confusion about the authorities' requirements. Without having a responsible office to regularly collect and provide easily accessible information about the availability of the flexibilities, DOD officials may not be fully leveraging them to further support the department's R&D goals.

DOD Used Selected Flexibilities to Support R&D Efforts but Faced Some Challenges

DOD's use of selected financial flexibilities from fiscal years 2017 through 2021 supported thousands of activities contributing to DOD R&D and efforts to modernize or innovate capabilities for military departments. The use of selected financial flexibilities varied and depended on several factors, such as having to meet specific criteria to use the flexibility or availability of funds. We found planning, guidance, and institutional support enabled DOD's use of the selected flexibilities, but DOD faced challenges when using some of these flexibilities. DOD officials cited numerous benefits that resulted from the use of selected financial flexibilities, including the ability to address R&D and operational needs or requirements that arise outside of DOD's planning, programming, and budgeting process.

DOD's Use of Selected Flexibilities Depended on Availability of Funding and Eligibility Requirements That Aligned with Needs

DOD reported making about $4.5 billion available from fiscal years 2017 through 2021 for the five selected financial flexibilities we reviewed to address specific lab needs, support technology development, develop and field capabilities to address specific threats, and fund software development (see table 3). This amount constituted less than half of the total amount allowed by the selected flexibilities from fiscal years 2017 through 2021, and constituted a small percentage of DOD's RDT&E appropriations overall.

Table 3: Reported Amounts Available for Selected Flexibilities from Fiscal Years 2017 through 2021

Dollars represent amounts DOD reported as available, rounded to nearest million. BA-8 was not available (NA) in fiscal years 2017 through 2020.

Flexibility	2017	2018	2019	2020	2021	Total
Funding Laboratory Enhancements Across Four Categories (FLEX-4)	299	459	530	559	620	**2,467**
Defense Research and Development Rapid Innovation Program, also known as Defense Rapid Innovation Fund (RIF)	250	250	241	-	-	**741**
Rapid Acquisition Authority (RAA)	424	-	155	18	-	**597**
Software and Digital Technology Pilot Programs, also known as Budget Activity Eight (BA-8)	NA	NA	NA	NA	588	**588**
Defense Laboratory Modernization Program (Lab Modernization)	-	-	-	111	-	**111**
Total	**973**	**709**	**926**	**688**	**1,208**	**4,504**

Source: GAO analysis of Department of Defense (DOD) information. | GAO-23-105822

DOD's use of the five selected flexibilities varied, in part, based on the availability of funding and the needs the flexibilities were designed to address. For example, FLEX-4 allows lab directors to use between 2 to 4 percent of their labs' available funds in support of activities in four categories that generally align with routine lab performance. In December 2018, we reported that this affords lab directors greater ability to make their own decisions over which activities the lab prioritizes and the means to fund those activities. In contrast, DOD used RAA as needed to meet specific urgent or emergent requirements to eliminate deficiencies that could result in the loss of life or mission failure, which led to more sporadic use during this time. The following further details use and conditions under which the funding can be used for each selected flexibility.

FLEX-4. FLEX-4 was the most frequently used of the selected flexibilities, funding thousands of activities across military department labs during fiscal years 2017 through 2021, according to DOD.28 Its overall use has increased across the military departments since 2017, when we previously found that the military departments were not maximizing their use of the flexibility.29 Some labs applied the full 4 percent allowed by statute to FLEX-4 activities. Other labs increased their use as of fiscal year 2022 or have plans to do so in the near future. According to DOD, each of the military departments takes its own approach to funding FLEX- 4.

DOD reported that FLEX-4 provides labs with flexibility to exploit scientific advances, respond to threats outside the PPBE cycle, and address lab- identified priorities. According to DOD, it provides funding for critical activities that would not otherwise receive funding. For example, a quarter of the Air Force's FLEX-4 basic and applied research category supports seedling initiatives to prove new concepts—providing initial funding for initiatives that could contribute to key future advances, according to DOD. Air Force officials explained that without FLEX-4 spending minimums, lab funds may be redirected to technologies with existing missions. FLEX-4 also helps by offering support for building and shaping labs' talent pool in new and emerging technology areas, according to DOD. For example, officials from some of the selected activities said FLEX-4 offered opportunities to grow and deepen staff knowledge and experience in the areas of artificial intelligence and autonomy.

RIF. Congress directly funded RIF in fiscal years 2017 through 2019, supporting hundreds of R&D and technology demonstration activities across DOD. However, Congress has not appropriated funding for the program since fiscal year 2019, and DOD did not include RIF in its fiscal year 2020 or fiscal year 2021 RDT&E budget requests. RIF program officials said that DOD uses RIF's original appropriation and provides funds in response to purchase requests based on a project or administrative request, rather than leveraging the flexibility's authority to transfer funds to the RDT&E account of a military department or unified combatant command for special operations forces. Officials said that this approach gives DOD RIF officials more control to reallocate funds across activities. For example, funds may become available for reallocation if activity costs are lower than expected or an underperforming activity is terminated.

Government and industry, small businesses in particular, have raised concerns over the lack of funds to mature technologies enough to be included in an acquisition program or delivered to the warfighter. Military department and OUSD(R&E) Manufacturing Technology officials said that RIF provides such funds and, without RIF, the selected activities would have been delayed or otherwise unsupported. As of July 2022, RIF officials said that 50 percent of activities funded using fiscal years 2017 through 2019 appropriations have transitioned or have plans to transition to operational use. A couple of RIF program managers said that the 50 percent transition rate means they are taking on appropriate risk to achieve innovation, and a higher transition rate would mean that they are not investing in new technologies.

RAA. RAA's use varied from fiscal years 2017 through 2021 because it is used in limited circumstances, as urgent needs generally arise outside of the normal PPBE cycle. Some officials have called RAA "a last resort" because it is used when immediate action is needed and when no other funding source is available. DOD reported using RAA a total of 13 times in fiscal years 2017, 2019, and 2020, each year staying below the limits allowed for each category annually. RAA users needed to identify funds from any existing DOD appropriations to acquire available solutions or products requiring minimal development to fulfill the urgent requirements. For example, a Marine Corps official said that they identified unused O&M dollars from a lower priority activity to purchase an available uncrewed aircraft system from industry to address an urgent operational need.

RAA users said that other funding mechanisms, such as reprogramming, could support urgent or emerging needs. However, officials we spoke with said that other funding mechanisms can take too long to execute and solutions risk becoming irrelevant when addressing immediate needs.

BA-8. DOD received fiscal year 2021 RDT&E funds for eight software development programs in the pilot program. These participating programs represented several departments across DOD and varied in size. DOD reported fiscal year 2021 funding for participating programs ranged from approximately $11 million at some departments to $230 million at others. DOD's internal selection criteria included that nominated programs had to previously have been fully-funded and preference was given to programs already participating in separate, Agile-related pilot programs. DOD proposed adding other programs to the BA-8 pilot in its subsequent budget requests. However, according to the report accompanying fiscal year 2023 defense appropriations, the appropriation committees' agreement encouraged DOD to stop proposing additional programs until it first demonstrated its ability to collect quantitative data on performance improvements provided by the pilot program.

According to DOD, effective software engineering typically requires concurrent technical work addressing bug fixes and existing vulnerabilities while developing new capabilities. These tasks may map to different appropriation categories based on statute and DOD financial regulations. However, BA-8 allows approved programs to use RDT&E funds for tasks that might otherwise be covered under multiple, separate appropriation categories. According to OUSD(A&S), BA-8 is not viewed as a "silver bullet." While it helps address some challenges for adopting commercial software development practices, it will not resolve all issues. OUSD(A&S) officials explained that a program office can use multiple appropriation categories when developing software using an Agile approach, but the flexibility to use one appropriation category can make it easier.

Lab Modernization. Lab Modernization was the least used of the selected flexibilities, funding three Air Force construction activities in fiscal year 2020. DOD requested and received $111 million of the maximum $150 million that the flexibility allows in any fiscal year. Its use, similar to RAA, is at DOD's discretion. DOD must include Lab Modernization military construction projects in the annual budget submission to Congress. Users of this flexibility must adhere to MILCON planning and reporting procedures, such as completing a planning and estimate document included in DOD's request for construction funding.

Some officials we spoke with expressed concerns about this flexibility, indicating that infrequent use could be due to funding procedures and noted that requests to use this flexibility could negatively affect labs' funding. For example, a request to use RDT&E funding for a construction project that otherwise would use MILCON funding could give the impression that a lab does not need the RDT&E funding for its non- construction R&D efforts. In addition, some officials said that there was confusion because, when the Air Force used the flexibility, the funds were provided using MILCON instead of RDT&E funding. Further, in a report, he Senate Appropriations Committee stated that it supported the activities DOD proposed using the flexibility and understood DOD's challenge in prioritizing small but critical lab construction projects. However, it encouraged DOD to request MILCON funds rather than RDT&E funds as allowed by the flexibility. According to Air Force officials, the department's MILCON approval processes take too long to meet high-priority RDT&E construction needs. For example, Air Force officials said that it could take between 5 to 15 years to get a project through the military department's approval process.

Planning, Guidance, and Institutional Support Enabled DOD's Use of Flexibilities, But Users Faced Challenges with Some Flexibilities

Based on our analysis of interviews with users of the five selected financial flexibilities, we identified three factors—(1) planning, (2) guidance, and (3) institutional support—that enabled effective use of the flexibilities.

Planning. This factor refers to actions that DOD officials took prior to using a selected flexibility. DOD and military department officials described planning as critical to leveraging each of the five flexibilities. Specifically, planning helped officials align flexibility activities with agency priorities, structure activities to meet desired outcomes, mitigate externalities hindering the use of flexibilities, and combine the selected financial flexibilities with other authorities, such as direct hire authority and other transaction authority, to optimize their use. For example, officials from all three military departments stated that planning helped align FLEX-4 minor military construction or repair of laboratory infrastructure and equipment activities with their modernization priorities. To that end, Army officials told us that they built the Robotics Research Collaboration Campus with FLEX-4 funding to provide expanded capabilities for the experimentation and testing of autonomous systems— a DOD modernization priority—at a more accessible location. Moreover, military department officials noted problems associated with delays in the availability of funding needed to initiate new projects using the financial flexibilities. Some of these officials stated that delays in using the new budget activity led to the program offices having to use an alternate approach while a continuing resolution was in effect and dealing with financial systems processes afterwards. Planning can help officials decide how to execute funding and structure their projects to accommodate such delays.

Guidance. This factor refers to the availability of formal documentation that specifies roles, responsibilities, and procedures for using a flexibility. DOD or military departments established guidance for four of the five selected flexibilities. Appendix III lists the primary guidance associated with each flexibility. There is no formal guidance governing Lab Modernization, and potential users of the Lab Modernization flexibility told us that they were unsure how to use it. For example, Air Force Research Laboratory officials said that they did not use this flexibility, in part, because of difficulty in understanding how to use it. Air Force Test Center officials used the flexibility but said the lack of guidance made obtaining approvals more difficult. Specifically, a Test Center official said that they had to educate staff in numerous other DOD organizations each time the Test Center attempted to use the flexibility. Further, guidance could clarify for officials when to use this flexibility in-lieu of requesting MILCON. An official within OUSD(R&E) told us that they informed potential users of this flexibility in the past but did not provide guidance. They said that the language of the flexibility is self-explanatory and they had not received requests for clarification. However, the agency is responsible for identifying departmental procedures for using the flexibility, such as the organizations responsible for approving its use. After we brought the lack of guidance to OUSD(R&E)'s attention, an official said that they plan to issue policy for Lab Modernization in fiscal year 2023.

Standards for Internal Control in the Federal Government states that management should communicate quality information down and across reporting lines to enable personnel to perform key roles to achieve objectives, address risks, and support the internal control system. Our past work has also recommended DOD develop guidance for using flexibilities, such as RAA, which it did. Guidance could facilitate DOD's use of Lab Modernization to expedite construction efforts in accordance with this authority and address any questions about approvals or the flexibility's relationship with MILCON funding.

Institutional support. This factor refers to having organizational leaders or officials who work directly with programs using the flexibilities advocate or provide the management and organizational infrastructure to facilitate their use. DOD and military department officials using the five selected flexibilities described institutional support as an enabling factor.

- Advocacy. DOD and military department leaders have demonstrated support through consistent, and in some cases, increased, resources for some flexibilities. Air Force, Army, Navy, and OUSD(R&E) officials described FLEX-4 as critical to DOD's modernization and technological advances. For example, Navy officials stated that FLEX- 4 fostered collaboration between experts in modeling and simulation as well as artificial intelligence to learn how coordination of autonomous vehicles perform in a variety of tactical scenarios. With BA-8, DOD leaders proposed the flexibility to Congress in DOD's Fiscal Year 2021 budget request and, in 2021, DOD received authority to pilot eight BA-8 programs. DOD requested pilot expansion in subsequent years that could help DOD better understand the use of BA-8 and knowledge acquired across different software programs that used the flexibility.

- Management and organizational infrastructure. According to multiple users, the Joint Rapid Acquisition Cell within DOD played an important role in facilitating the military departments' use of RAA. For example, a Marine Corps official said that Joint Rapid Acquisition Cell support staff was helpful in moving the Marine Corps' requirement through the RAA process and ensured that the RAA package was appropriately staffed. FLEX-4 users across the Air Force, Army, and Navy also described knowledgeable officials within the labs they could turn to with questions about the process for proposing activities for FLEX-4 funding and for support when using the flexibility. At the local level, an official at Edwards Air Force Base advocated for the Lab Modernization flexibility and, despite the lack of guidance, developed procedures for using it to support test center construction at three locations.

By comparison, we identified a lack of institutional support in the RIF program. For example, DOD did not include RIF in its Fiscal Year 2020 or Fiscal Year 2021 RDT&E budget requests, and Congress did not appropriate funding for it. A DOD official stated that DOD senior leadership did not support RIF as a funding priority at that time but anticipated DOD leadership may include RIF in future budget requests.

The official explained that previously, leadership may not have fully understood the importance of this program and its effect on the science and technology community, in part, due to the RIF program's lack of reporting on its work. DOD has updated the RIF implementing procedures to emphasize connections to DOD's modernization priorities and identify an office responsible for the program. DOD has also enhanced its guidance and reporting, and developed its organizational infrastructure for reviewing proposals and making awards with an aim of shortening its timelines.

Some officials said that they encountered resistance when using BA-8 and RAA flexibilities because these require deviation from the execution of funding that officials were accustomed to using. For example, a Space Force official using BA-8 said that both experienced and junior financial management staff were hesitant to use the RDT&E budget activity for sustainment or procurement activities because the flexibility goes against established procedures or they were unfamiliar with what the flexibility allowed. In contrast, institutional support helped address resistance that could discourage or slow flexibility use. For example, a DOD official working on another BA-8 program described an environment in which the entire program office, including financial management staff, were committed to making this flexibility work. They said that staff acquired expertise and familiarity with what the flexibility allowed, helping them to maximize benefits of BA-8.

DOD Used Selected Flexibilities to Accelerate Funds to R&D Efforts

According to the users of the five flexibilities who we interviewed, the flexibilities' use allowed them to address R&D and emerging needs more quickly by avoiding certain steps in the PPBE process. Agency officials stated that, because the PPBE process can take several years to make funds available for use, innovation opportunities or emerging needs can be difficult to address.

The flexibilities supported users' efforts to revitalize or refurbish labs and test centers, begin early research, mature technologies to transition into programs, and promote software development, among other things.

Agency officials who used the flexibilities said that without them, their projects would have experienced delays, delivered less capability to address a need, or run the risk of being unfunded.

Below are high-level summaries of each selected flexibility, including DOD identified benefits; the PPBE phases streamlined; and examples of the contributions to research, development, innovation, and modernization.

Overall, DOD and agency officials said that FLEX-4 contributes to innovation, the military departments' modernization, and national defense strategy by expanding knowledge. Officials said FLEX-4 also increases the capacity and size of the workforce and creates opportunities to explore, develop, and test new technologies and their potential uses. Further, it streamlines parts of the PPBE process (see fig. 3).

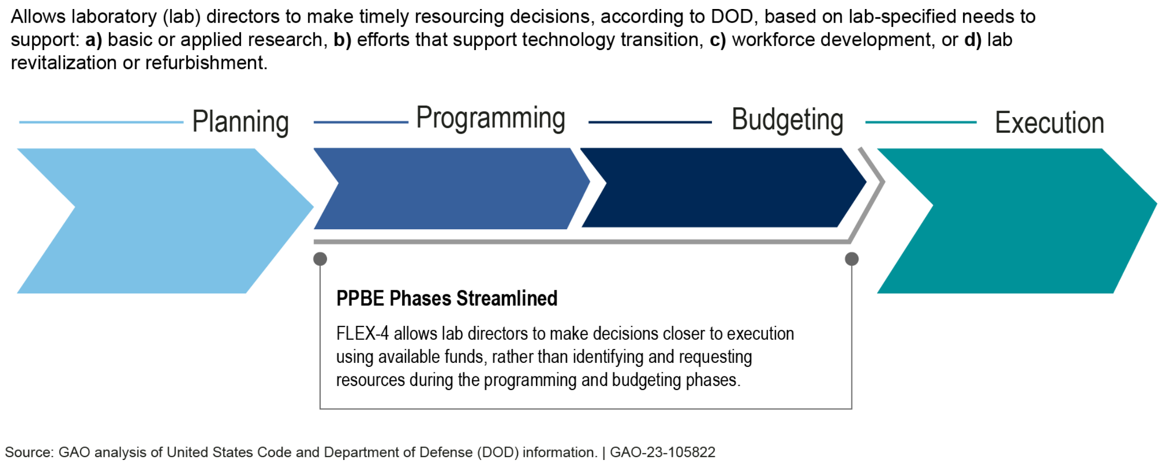


Figure 3: Funding Laboratory Enhancements Across Four Categories (FLEX-4) Benefit and Planning, Programming, Budgeting, and Execution (PPBE) Phases

Specific examples officials identified include:

- **Expanded research and testing opportunities.** Air Force, Army, and Navy officials said that FLEX-4 positions labs to conduct current and future research and testing. For example, Air Force officials for the Enriched Understanding of Hypersonic Materials activity said that the flexibility is supporting hypersonic material testing and simulation efforts. They are testing materials and developing prediction models that will help inform the next generation of materials. Without FLEX-4, officials said that the activity would be delayed several fiscal years.

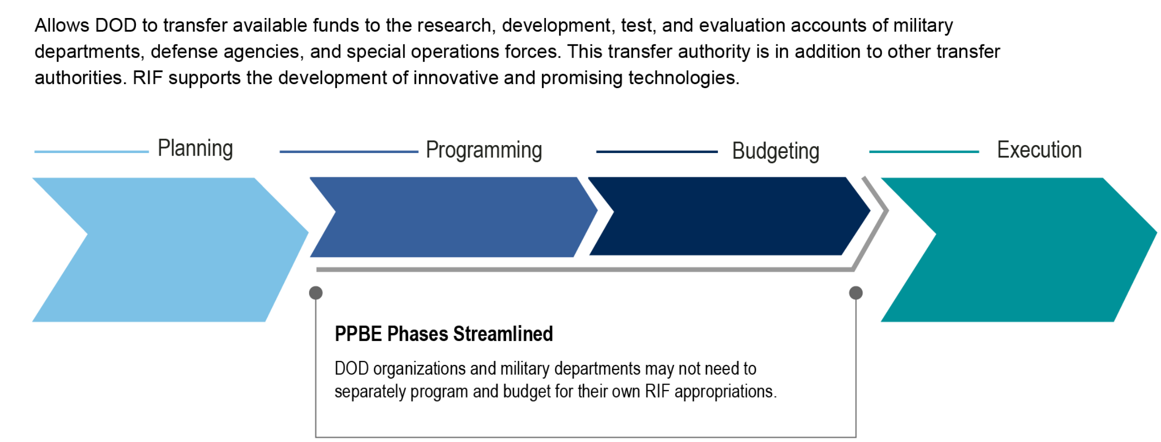
- **Workforce development opportunities.** Navy and Army officials said that FLEX-4 increased workforce development opportunities. For example, Army officials said that the Distinguished Postdoctoral Fellowship and Research Associateship Program helps bring in top- level scientists and engineers to better address the Army's innovation and modernization needs. These participants can introduce new techniques to a lab and expand lab relationships with universities.

- **Seed funding for early research**. Army and Air Force officials said that FLEX-4 provided funding for future efforts. For example, the Army used FLEX-4 funding to jump-start its Emerging Overmatch Technology activity. Army officials said that the flexibility was critical in maturing the technology and demonstrating the uncrewed aircraft systems' ability to achieve cooperative protection for small units of combat vehicles. As a result, they said that the Army has requested funding through the PPBE process to further develop this technology.

- **Investment in lab infrastructure.** Air Force, Army, and Navy officials said that FLEX-4 provided critical funding for lab infrastructure, ranging from investing in equipment to refurbishing and renovating buildings. Air Force officials said that FLEX-4 is meant to help the labs keep pace with some infrastructure needs, despite what they view as a lack of prioritization for DOD R&D infrastructure, which they said is a strategic issue for the department. However, officials from each of the military departments said that the $6 million cap on minor military construction limits the types of investments labs can make for repair or minor military construction of laboratory infrastructure and equipment. Since we spoke with these officials, Congress increased the cap to $9 million.

- **Increased collaboration with program offices, within the labs, and with outside entities.** Army and Navy officials said that their use of FLEX-4 provided opportunities for collaboration within their military departments and with industry. The Army used FLEX-4 funds to construct facilities with convenient and collaborative spaces, and the Navy used FLEX-4 funds to support cooperative agreements with industry. In both situations, agency relationships benefitted from the availability of the funds, resulting in time and cost savings. For example, Ship-to-Shore Navy officials said that they collaborated with other warfare centers and industry partners to develop a water-based, small, uncrewed surface vehicle—which served as a proof of concept for similar technologies. Navy officials said that industry partnerships provided additional expertise and the prototype vehicle used. Further, the team received important feedback from potential users following the demonstrations, which we have previously identified as a leading practice when developing new technologies.

The RIF flexibility allows DOD to transfer available funding to expedite support for further developing technologies that solve operational challenges and contribute to addressing national security needs. RIF funded awards that aim to transition technologies to military programs. It offers opportunities to streamline the PPBE process (see fig. 4).



Source: GAO analysis of United States Code and Department of Defense (DOD) information. | GAO-23-105822

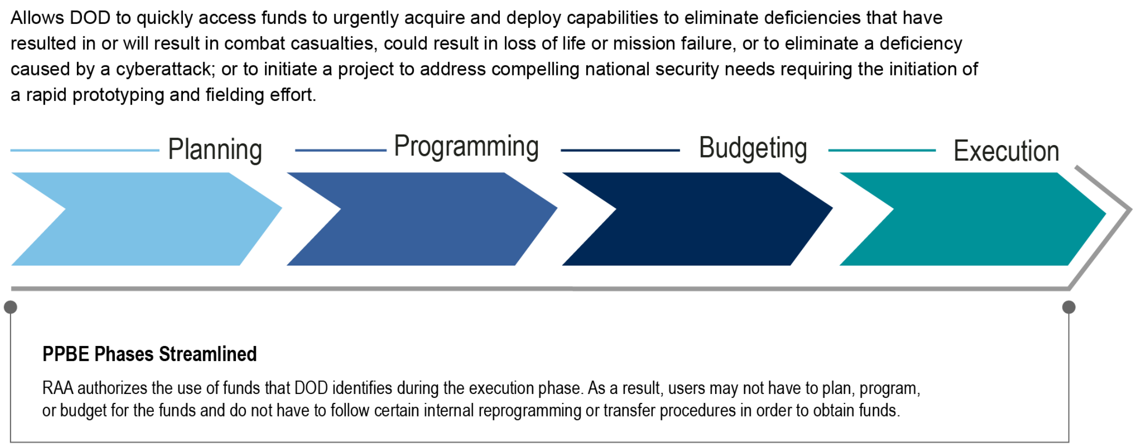
Figure 4: Defense Research and Development Rapid Innovation Program, also known as Defense Rapid Innovation Fund (RIF) Benefit and Planning, Programming, Budgeting, and Execution (PPBE) Phases

Specific examples of benefits officials identified include.

- **Assisted technology transition.** OUSD(R&E) and Navy officials said that the RIF program provides funding to bridge experimental research and acquisition programs. For example, a Navy official said that, when the Rapid Acquisition Sensor and Response activity first received RIF funding, the technology to track submarines was at the early lab development phase, but has since moved to operational environment testing. They said that the activity's technology now has a program office to sponsor its transition into a program of record. The Navy official said that the maturation of the technology or the interest in the activity by a program of record would not have been possible without RIF support.

- **Informed future strategies.** An Army official said that RIF activities that do not transition to a program of record can help inform future efforts. For example, they said that the results from the Mobile Ad-Hoc Networking in Congested and Contested Environments Prototype's activity assessment, which included potential users, provided valuable information for shaping other network design goals. Additionally, the technology remains a consideration for future communication capabilities.

According to DOD officials, RAA is beneficial in cases where there are insufficient resources to address an urgent need, such as preventing loss of life. It also streamlines parts of the PPBE process (see fig. 5).



Source: GAO analysis of relevant defense authorization acts and Department of Defense (DOD) information and interviews. | GAO-23-105822

Figure 5: Rapid Acquisition Authority (RAA) Benefit and Planning, Programming, Budgeting, and Execution (PPBE) Phases

Specific examples of benefits officials identified include:

- **Reduced internal barriers in meeting urgent or emergent needs.** Air Force, Army, and Marine Corps officials said that they could have addressed certain needs without RAA, but officials would not have been able to obtain the solution quickly enough, or at the speed of relevance, to meet their urgent or emergent needs. For example, Army officials for the Counter-Small Unmanned Aircraft Systems activity said that they could have used reprogramming, but to do so would have taken 1 to 2 years longer.

- **Solutions met immediate needs and provided enduring capability.** Army, Marine Corps, and Air Force officials said that there were limited solutions available to address different emerging needs, which resulted in using or building on existing solutions in the commercial sector to develop or procure a new capability that could be used in other situations going forward. For example, Air Force officials said that the COVID-19 airlift activity not only met the urgent need for transporting COVID-19 patients while keeping the crew safe, it is available to transport patients with other deadly diseases.

Officials said that benefits of BA-8 are primarily related to time and labor savings by staff spending less time on administrative activities, such as programming and budgeting for multiple appropriations. BA-8 streamlines parts of the PPBE process (see fig. 6).

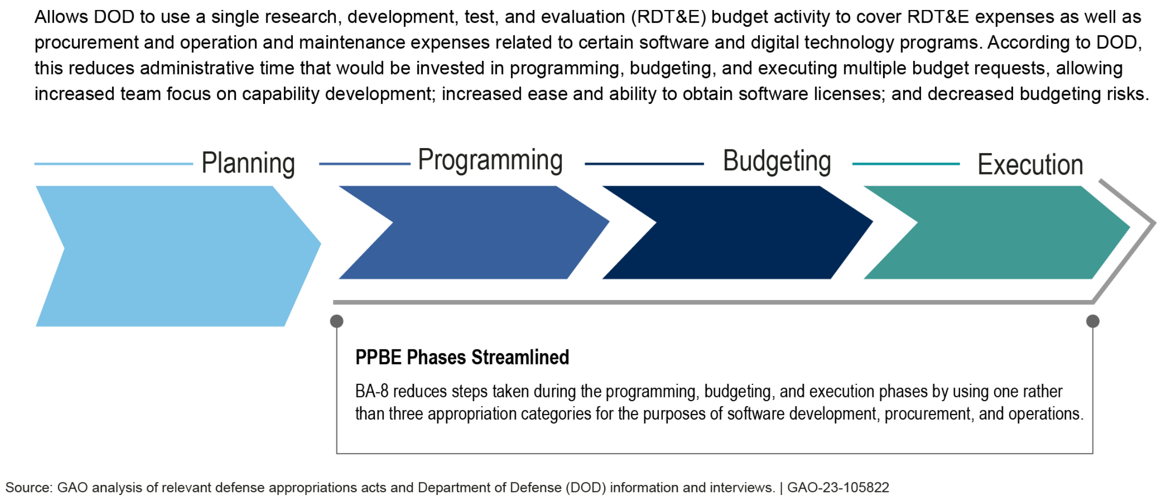


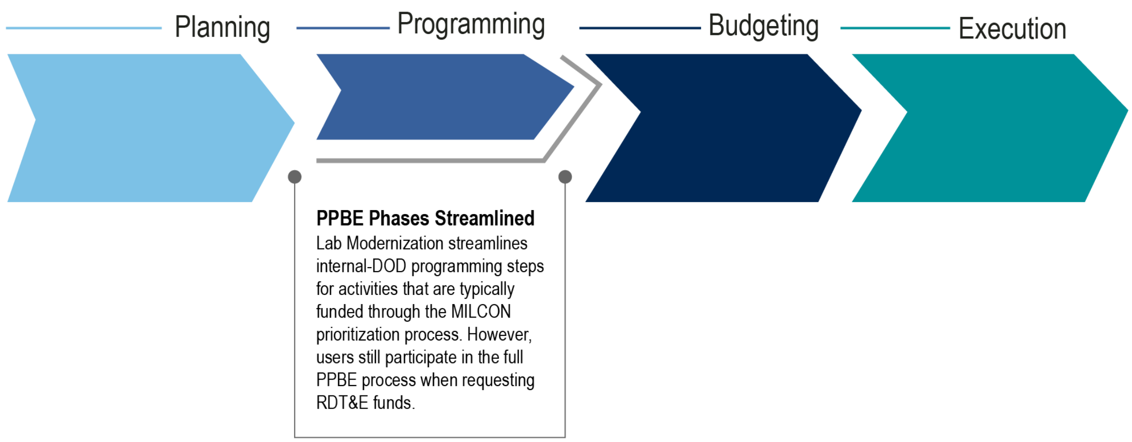
Figure 6: Software and Digital Technology Pilot Programs, also known as Budget Activity Eight (BA-8) Benefit and Planning, Programming, Budgeting, and Execution (PPBE) Phases

Specific examples officials identified include:

- Increased focus on developing capabilities. Space Force and Army program officials said that BA-8 allowed their teams to focus more on providing capabilities to users, such as tools to detect cyber threats on department networks, rather than on activities or steps that occur during the programming or budgeting PPBE phases. For example, Space Force program officials said that BA-8 allows them to operate with lean financial management staff focusing more resources on the technical aspects of the program.
- Solutions that better align with customer and capability needs. Army and Space Force program officials said that BA-8 helped when program requirements shifted from needing to buy a renewal license to purchasing new software. Without flexible funds, these officials said that they would have likely selected a solution based on available funding options rather than using BA-8's available funding to find a solution that best met program needs.
- Reduced budget risk for program offices. Navy program officials said that they would not have accurately predicted RDT&E needs when creating their BA-8 program's fiscal year 2022 budget request. During execution, they told us that they learned the software needed significantly more development than previously expected. Without the flexibility offered by BA-8 to pivot between development and maintenance efforts, officials said that they would have delivered less capability in fiscal year 2022.

Lab Modernization can allow labs and test centers to build or rehabilitate facilities to operate using the latest technology. It also streamlines part of the PPBE process (see fig. 7).

Allows DOD to use research, development, test, and evaluation (RDT&E), rather than military constructions (MILCON), funds for the purposes of building or revitalizing certain facilities, such as labs and test centers. By using RDT&E appropriations, RDT&E decisionmakers compare proposed activities against RDT&E needs thereby reportedly avoiding competition with other MILCON activities and the resulting, lengthy programming process.



Source: GAO analysis of relevant defense authorization and appropriations acts and Department of Defense (DOD) information and interviews. | GAO-23-105822

Figure 7: Defense Laboratory Modernization Program (Lab Modernization) Benefit and Planning, Programming, Budgeting, and Execution (PPBE) Phases

Officials identified that Lab Modernization also:

- Provided a funding path for lab and test centers that would otherwise not be available. Air Force officials said that RDT&E projects do not compete well in the MILCON process or rate highly on the prioritization list. They explained that construction addressing the health and safety of the service members and their families are higher priorities in the budget process. The Air Force used Lab Modernization to construct a Joint Simulation Environment facility at Edwards Air Force Base. This facility provides new testing capabilities for the F-35 Lightning II and other aircraft for entities across DOD. Air Force officials said that, without Lab Modernization, the construction of the facility would not be possible as there would not have been a funding path to support it.

BA-8 Financial Flexibility Partially Met Leading Practices

We found DOD partially met leading practices we identified in prior work related to pilot design for BA-8, the one selected flexibility that is currently a pilot program. Our previous work found that implementing these leading practices for pilot design can help ensure agency assessments of the pilot produce the information needed to make effective program and policy decisions. Such a process enhances the quality, credibility, and usefulness of evaluations, in addition to ensuring time and resources are used effectively. The five leading practices form a framework for effective pilot design and evaluation. Figure 8 summarizes these five leading practices for pilot design.

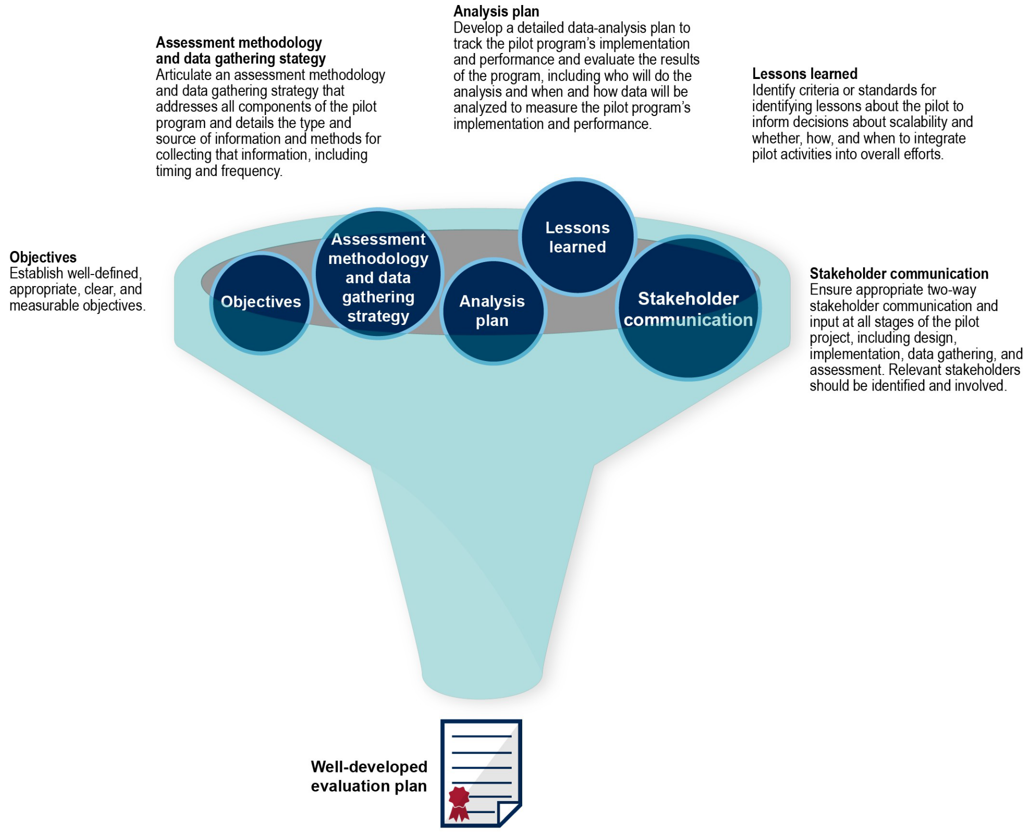


Figure 8: Leading Practices that GAO Identified for Pilot Program Design

Congress established the BA-8 pilot in fiscal year 2021. The pilot identified eight software programs allowed to use the single RDT&E budget activity and eight programs using the traditional appropriation categories—RDT&E, Procurement, and O&M—for comparison. Before implementing the pilot, DOD had to establish metrics and develop a plan for assessing each program using the single appropriation, such as comparing program performance against their own historical performance and a comparison group of eight, traditionally funded programs. DOD was directed to report quarterly on the pilot's progress.

DOD submitted a plan to congressional defense committees for assessing the pilot and developed guidance for implementing the pilot. However, we found that DOD has not fully met the five leading practices for pilot design (see table 4).

Table 4: Software and Digital Technology Pilot Program (BA-8) Partially Met Leading Practices for Pilot Design

Leading practice	GAO assessment (Met, Partially Met, or Did not Meet)
Establish well-defined, appropriate, clear, and measurable objectives	Met
Clearly articulate assessment methodology and data gathering strategies	Partially Met
Document lessons learned	Partially Met
Develop plan to evaluate pilot results	Partially Met
Ensure appropriate stakeholder communication	Partially Met

Source: GAO analysis of Department of Defense information. | GAO-23-105822

Establish well-defined, appropriate, clear, and measurable objectives. OUSD(A&S) is responsible for the BA-8 pilot and established measurable objectives in its implementation plan, such as BA-8's effect on programs using the single appropriation category. Although OUSD(A&S) officials said that they are adjusting some strategies described in their implementation plan, they expect BA-8's objectives to remain unchanged.

Clearly articulate assessment methodology and data gathering strategies. As of April 2023, OUSD(A&S) officials have not updated their assessment and data collection methodologies. At the start of the BA-8 pilot, OUSD(A&S) outlined metrics for participating programs, including descriptions, frequency, and methods of data collection, in its June 2021 implementation plan and pilot agreements. DOD used these agreements to ensure that mechanisms were in place to provide consistent monitoring and data collection for the pilot. However, the pilot ran for about a year and a half without all programs having agreed to ensure mechanisms were in place. Most participating programs signed their respective pilot agreements in fiscal years 2020 and 2021, but several programs using the traditional funding structure did not fully sign their pilot agreements until early fiscal year 2023. Furthermore, OUSD(A&S) officials said that they have not implemented the methodologies described in those documents. For example, OUSD(A&S) officials said that they realized metrics outlined in the pilot agreements, such as product development time and lead-time for changes, were not consistently measurable across participating programs. OUSD(A&S) officials said that they are in the process of establishing new strategies to better understand BA-8's effect on Agile software development for participating programs.

Document lessons learned. OUSD(A&S) officials said that they generally capture program feedback in a shared drive and ask programs to provide information about their experience. But they have not formally documented lessons learned and do not have plans in place to review lessons if any are collected. DOD plans to share lessons gathered during the pilot in its final report and identified program managers or their designee as being responsible for collecting them. Officials for the four selected BA-8 programs in our review told us that they share insights regarding their use of the new budget activity and participation in the pilot with OUSD(A&S) officials at monthly and quarterly meetings but do not formally collect lessons learned.

Develop plan to evaluate pilot results. OUSD(A&S) detailed its plans to assess the pilot in its June 2021 implementation plan. However, officials told us that they are not consistently collecting the data that would be used in their evaluation and, after consulting with participating programs, cannot use the planned metrics to evaluate the pilot. OUSD(A&S) officials said that they have yet to fully establish or document an updated evaluation plan.

Ensure appropriate two-way stakeholder communication. OUSD(A&S) officials communicate with participating programs on a quarterly and monthly basis, but officials said that affected programs were not involved in the design phase of the BA-8 pilot. OUSD(A&S) officials said that engaging stakeholders during BA-8's development might have helped them avoid using metrics that were not applicable for some programs.

In addition, OUSD(A&S) is in the process of responding to congressional concerns about the adequacy of its required quarterly reporting.

OUSD(A&S) is required to provide updates on the pilot's progress, but OUSD(A&S) officials said that the reports they provided were anecdotal and did not clearly address the Senate Appropriations Committee's requests for quantitative data. OUSD(A&S) officials said that members of Congress raised concerns about OUSD(A&S)'s reporting in November 2021 and did not want to add more programs without data to support claims about the pilot's effect. Further, the explanatory statement accompanying DOD's fiscal year 2023 appropriations encouraged DOD to refrain from submitting additional pilot programs until DOD can demonstrate its ability to provide data on performance improvements. OUSD(A&S) officials told us that they are working through potential solutions but have yet to establish new procedures or plans to analyze collected data. As OUSD(A&S) continues to adjust its strategies, incorporating the elements of a well-developed evaluation plan would better position DOD to provide more informative updates to Congress regarding the effectiveness of the pilot.

DOD has an opportunity to build on knowledge obtained over the past 2 years through interactions with stakeholders and address congressional concerns by using the leading practices for pilot design. Without a well- developed evaluation plan, including strategies for assessing lessons learned and BA-8's effect on participating programs, DOD and Congress will lack the information needed to assess the effectiveness of the pilot and whether this financial flexibility should be made permanent.

Conclusions

With research and development efforts, innovation and technology evolution can stem from bursts of sporadic insight that occur outside the PPBE cycle. DOD seeks to quickly identify and pursue promising emerging technologies for its innovation and modernization purposes, in part, by leveraging opportunities to streamline its lengthy PPBE process. Congress has provided a set of flexibilities to help DOD be more agile; however, the department could do more to take full advantage of them. DOD could use our work as a starting point for regularly communicating and disseminating information about the most recently available flexibilities and provide regular updates on any changes Congress may make to existing or new flexibilities. With easily-accessed information available department-wide, DOD would be better positioned to identify opportunities to leverage the flexibilities and the value they provide. In addition, having guidance on how to use these flexibilities could help DOD maximize their use. Furthermore, for pilot programs, implementing a well-developed evaluation plan can help DOD know what effect changes from its normal operations are having, whether they are generating the anticipated benefits, and whether there is a good business case to make the changes permanent.

Recommendations for Executive Action

We are making the following three recommendations to the Secretary of Defense:

The Secretary of Defense should ensure the Deputy Secretary of Defense designates a primary office responsible to regularly collect current information about the financial flexibilities that are available to support DOD's research and development, innovation, and modernization efforts and ensures the office makes the information easily accessible department-wide. (Recommendation 1)

The Secretary of Defense should ensure the Under Secretary of Defense for Research & Engineering develops guidance for the Defense Research Laboratory Modernization program that communicates the purpose, roles and responsibilities, time frames, procedures, and other relevant information needed to effectively implement this flexibility. (Recommendation 2)

The Secretary of Defense should ensure the Under Secretary of Defense for Acquisition & Sustainment implements an evaluation plan, developed using leading practices for pilot design for assessing the effectiveness of the Software and Digital Technology Pilot Programs, also known as Budget Activity Eight (BA-8). (Recommendation 3)

PPBE, Technology Transition, and “The Valley of Death”

Olivia Letts—is a Research Manager with the Baroni Center for Government Contracting. Letts’ research interests include defense spending and management, emergent dense technologies, and U.S. military and foreign policy. Prior to working for the Center, Letts worked as an Associate Analyst to support innovative companies entering the defense market and government contracting spaces. Letts received a BA from the University of Florida and an MA from Georgetown University in security studies. Her writing has been featured in Defense News, Army Magazine, The Cipher Brief, Breaking Defense, and the Georgetown Security Studies Review. [oletts@gmu.edu]

Edward Hyatt, Ph.D—is a Senior Research Fellow at the Baroni Center for Government Contracting in the Costello College of Business at George Mason University. He has a decade of research experience and another seven years of managerial experience in the public procurement and contracts profession. He holds several advanced degrees, including a PhD in business management from the University of Melbourne, and has participated in the research and publication process on dozens of projects. He has two streams of research, one centered on organizational behavior topics like personnel selection and managerial decision-making, and the other involving government acquisition and policymaking matters. [ehyatt4@gmu.edu]

Jerry McGinn, Ph.D—is the Executive Director of the Greg and Camille Baroni Center for Government Contracting in the School of Business at George Mason University (GMU). Prior to establishing the Center, McGinn served as the senior career official in the Office of Manufacturing and Industrial Base Policy in the Department of Defense. McGinn also served in the DoD as Special Assistant to the Principal Deputy Under Secretary (Policy) and as a political scientist at RAND. His writing has also been featured in multiple GMU and RAND reports, plus articles in various defense news outlets. McGinn was commissioned into the U.S. Army, served with distinction as an infantry officer, and is a graduate of Ranger and Airborne Schools. He has earned a PhD, MS, and MA from Georgetown University as well as a BS from the U.S. Military Academy. [jmcginn5@gmu.edu]

Abstract

Before they can be developed and deployed to help the U.S. warfighter accomplish its military objectives, all defense capabilities must obtain funding by passing through the stages of the Planning, Programming, Budgeting, and Execution (PPBE) process. While PPBE has undergone systematic changes since its inception in the early 1960s, various issues have been attributed to its largely unchanged framework. In particular, the defense community has reported PPBE-related setbacks affecting technology transitions, joint efforts, and program lifecycles. This paper explores six case studies for critical or cutting-edge defense programs and organizations and PPBE’s impact on their progress. Findings suggest PPBE can slow the development of new capabilities supporting the warfighter, hamper fiscal flexibility, and make it harder for programs to adjust to the evolving needs of the combatant commands and services. However, findings from the six case studies also suggest PPBE’s impacts on technology transition are often exaggerated by the defense community. PPBE-related challenges can also be mitigated through strong senior leadership, the consolidation of program elements, the use of agile approaches such as the Middle Tier of Acquisition, sufficient congressional engagement, and other special efforts.

Introduction

In support of the efforts of the Commission on Planning, Programming, Budgeting, and Execution Reform, George Mason University’s Baroni Center for Government Contracting was tasked with the following research objective:

Pursuant to Sec. 1004(f)(2)(c), conduct “a review of how the [PPBE] process supports joint efforts, capability and platform lifecycles, and transitioning technologies to production.”

To address this research task, the research team examined the role of the PPBE process within the context of six diverse U.S. Department of Defense (DoD) programs and organizations. The case studies comprised the following four programs and two organizations:

- Navy Large and Medium Unmanned Surface Vessels (LUSV/MUSV)
- Air Force Collaborative Combat Aircraft (CCA)
- Army Robotic Combat Vehicle (RCV)
- Army Tactical Intelligence Targeting Access Node (TITAN)
- The Space Development Agency (SDA)
- Joint Rapid Acquisition Cell (JRAC)

All six case studies were chosen for their current relevance, operational importance, or dependence on cutting-edge technologies to meet joint strategic or Service-specific needs. All case studies, with the exception of JRAC, also explored fairly new programs, in which the speed of technological development and program advancement have thus far been critical to enable early success. Each of the case studies were conducted using the following methodology. First, the research team conducted a literature review of key publicly available documents, including the DoD's budget justification books (J-books). After identifying major issues inherent to the PPBE process and its functions, the research team conducted interviews with over 20 subject matter experts associated with the programs and organizations. The majority of interview participants were key government personnel, but industry perspectives were also provided by personnel associated with contractors on several of the case study programs. The interviews abided by the Chatham House Rule whereby all identities of the interview participants and information during the interview are to remain unidentified.

Top-level case study research findings aligned with common PPBE criticisms. The research team observed that PPBE had tangible impacts on the rapid development and deployment of new capabilities to support warfighter needs and complicated the government's ability to accommodate adjustments needed to rapidly respond to evolving programs and requirements. The PPBE process posed added obstacles when the need for fiscal flexibility was greatest, particularly during the year of execution. However, many of the widespread PPBE criticisms reported by the defense community were found to be exaggerated, as case study interview participants cited a wide range of other exogenous factors affecting program success.

Research efforts focused on the link between PPBE and technology transition. The objective was to examine whether the PPBE process is a root cause of technology transition failure in the so-called "valley of death" as experimental projects evolve into programs of record. Therefore, as part of the case study reviews, the research team was also asked to address two crucial questions of technology transition, including:

1. Are higher-valued opportunities foregone at the expense of continuing lower-valued programs?
2. Is the PPBE process a significant root cause of failure to reallocate resources to higher-valued uses as distinct from the JCIDS or Small "A" acquisition process?

In response to the two questions of technology transition, the research team concluded the affirmative for both, but with caveat. Interviews with subject matter experts revealed that higher-valued opportunities are indeed delayed or foregone due to the PPBE Program Objective Memorandum (POM) cycle's tendency to prioritize pre-existing programs. In some cases, these pre-existing programs were considered of lower-value or not seeing adequate returns on funding. The PPBE process was cited as one cause of failure to reallocate resources to higher-valued uses, including urgent warfighter needs and cutting-edge programs. As with other top-level findings obtained by the research team, the PPBE's relationship to the valley of death was nevertheless found to be one of many other influencing factors—not necessarily the *root* cause.

The case study research findings were not all negative regarding the PPBE process, however. The success of certain defense programs seemed to suggest that cutting-edge defense capabilities can traverse the normally lengthy PPBE process without lagging in the valley of death. The PPBE process was found to have acted as less of a hurdle when programs were championed by strong senior leadership or entailed congressional engagements characterized by consistency and cooperation. The PPBE process was also found to be a more neutral influence in programs that utilized agile approaches such as the MTA pathway or broader program elements (PEs) in their budget structure enabling flexibility in program execution.

This paper seeks to provide in-depth yet concise summaries for the PPBE-related impacts on the technology transitions and program success within the six case studies and to present possible recommendations for PPBE reform or for navigating program success to defy limitations inherent to the PPBE process. Where possible, this paper also seeks to assess how the case study findings aligned with the recommendations explored by the PPBE Reform Commission in its interim and final reports. To set the context in which the research was conducted, the paper will provide a brief background of the PPBE process and the factors that led to the formation of the PPBE Reform Commission. The paper will then explain key findings for each case study. Lastly, it will provide relevant conclusions and recommendations.

Background

Planning, Programming, Budgeting, and Execution (PPBE) is a calendar-driven process used by the DoD to allocate resources in support of its capability needs.

In PPBE's efforts to align top-level, long-term strategy with optimal resource allocation, each step must conform to various fiscal, time-related, and other constraints. It can take 2–4 years for a defense program to transverse the PPBE process through each of the phases up to and through the contracting stage. The ability of a weapons system to transverse the PPBE process is also tightly linked to the other elements of the DoD's acquisition process trifecta, which also includes the Joint Capabilities Integration and Development System (JCIDS) for the defining requirements of a weapons system and the Defense Acquisition System (DAS) for guiding the multiple acquisition pathways of defense capabilities. PPBE was formerly known as PPBS, or the Planning, Programming, and Budgeting System. Former Secretary of Defense Robert McNamara established PPBS in the early 1960s with the aim of using scientific protocol and management methods to align the DoD's strategic needs with capabilities and reduce wasteful redundancies in the defense budget (Sapolsky et al., 2017).

Many entities have a stake in the budget formulation and oversight: three Service Secretaries, five Service Chiefs, the Office of Management and Budget (OMB), four congressional defense committees, and the defense industrial base comprising those companies under contract with the DoD. Each military department conducts the PPBE process slightly differently to generate and justify their shares of the budget. While the Executive Branch maintains its leverage as executor of the budget, it is still the weaker player vis-à-vis Congress, which maintains its Constitutional authority to "provide for the common Defense" (MacGregor et al., 2022). Appropriation accounts (e.g., RDT&E; Procurement) PEs and other organizational subdivisions within defense programs such as budget activities (BAs) comprise the main organizational features of defense programs within the J-books, compiled during the Budgeting phase to provide details to Congress that justify program budget requests.

Each phase of PPBE serves of a distinct role. The Planning phase identifies changes in the strategic environment and necessary updates for the military's strategic allocation, directing the ensuing Programming phase. During the Programming phase, the military services develop their own Program Objectives Memorandum (POM) outlining their intended resource allocation and priorities to meet their future objectives. This kickstarts the Program and Budget Review (PBR) cycle, in which OSD leadership and the DoD's Office of Cost Assessment and Program Evaluation (CAPE) evaluate the POMs. In the Budgeting phase, the defense budget is formalized within documentation that contains program cost estimates and complies with top-level strategy and regulation—as subjected to OSD and congressional review. The defense budget is obligated, and funding deployed, under a continuous program performance feedback loop, in Execution. The phases of PPBE often overlap; while one POM is being built, funding could be executed for several years' worth of prior year funding. Programmatic changes can occur up to the last minute in the Programming and Budgeting phases, but within top-line funding and oversight constraints set by OSD (*Interim Report*, 2023).

Case Study 1: Navy Large and Medium Unmanned Surface Vessels (LUSV/MUSV)

For its first case study, the research team examined two new U.S. Navy programs in tandem: the Large Unmanned Surface Vessels (LUSV) and Medium Unmanned Surface Vessels (MUSV) programs. The programs comprise two different variants of unmanned surface vessels (USV) to be developed, fielded, and remotely operated by the Navy in a semi-autonomous or fully autonomous fashion. The two USV variants are described in the FY 2024 Navy Budget Justification Book as "affordable, high endurance, reconfigurable ships able to accommodate various payloads for unmanned missions and augment Navy's manned surface force." LUSV and MUSV will support the Navy's transition from a traditional emphasis on fewer high-dollar ships toward a fleet with more low-cost and adaptable USVs (Zoldi, 2023). Both programs benefited from the progress of previous research programs, with MUSV inheriting Sea Hunter and Seahawk ships from the DARPA/Office of Naval Research Medium Displacement Unmanned Surface Vessel program, and LUSV inheriting four ships from the Overlord Unmanned Surface Vehicle development effort within the Ghost Fleet program run by the DoD's Strategic Capabilities Office/Uncrewed Maritime Systems Program Office (*Uncrewed Maritime Systems*, 2022).

LUSV and its associated USV Enabling Capabilities program were new starts in FY 2020. The Navy is currently using LUSV prototypes to develop new concepts of operations and intends to equip the reconfigurable model to carry multiple launch tubes for anti-ship and land-attack strike payloads (Zoldi, 2023). MUSV, identified in the Navy J-books as the Medium Displacement Unmanned Surface Vehicle (MDUSV) program, was a new start in FY 2019; it was designated in the FY 2024 Navy Justification Book as a Rapid Prototyping Program following the MTA pathway for rapid acquisition and delivery. Despite growing uncertainty about the utility of MUSV due to its surveillance mission overlap with smaller and cheaper programs, the Navy has moved forward with its plans to develop and procure the medium-sized variant (Eckstein, 2022).

LUSV/MUSV Program Key Finding #1: Several aspects of the PPBE process make it more cumbersome to move a program forward. Several LUSV/MUSV interview subjects expounded on inefficiencies when managing congressional marks during program and budget review. One interview subject observed that when the House Armed Services Committee and Senate Armed Services Committee marked two different amounts for one funding line spread across three programs versus individual programs within the line for the USV program, the Navy ultimately appealed to the slightly lower mark, but this rendered one program component inexecutable. Another interview participant observed that the USV program office struggles without being able to move monies in or out of marked-with-prejudice budget lines, reducing program flexibility.

The tight execution schedule of the PPBE was also cited as being highly incompatible with continuing resolutions, which puts program offices behind when it comes to executing their funds. Defense programs typically cannot access new funds until two months after a continuing resolution ends, resulting in only half a year of properly funding execution. According to interview subjects, this can foster perceptions of non-performance by Congress:

> In a normal [program], the PPBE should work. Congress passes October 1st, OSD can say, OK, this is the impact to 24 programs. I can adjust 25 to maybe so when we submit to Congress, you know in February things will work. The sponsor can look at that and say … I can adjust 26. We haven't been able to do that in probably a decade. So, any problem where you get a little bit behind … Congress marks you for being behind [even though] they start you six months late ... This is what you have to plan for or you're going to run out of money before the money shows up in the execution year.

In the LUSV/MUSV case study interview, there was also a brief discussion about budgeting constraints which limit a program's evolution, namely the small size of the reprogramming threshold limit and the lack of a management reserve. First, the interview subjects said that the reprogramming threshold limit was too low considering the overall size of budgets. They felt that the low limit hamstrings a responsible department from solving smaller issues on their own recognizance. As summarized by one of the interview subjects:

> The $10 million below threshold reprogramming versus the above threshold reprogramming and the need to go talk to all four committees about the moving money means that even if we did have a good idea of something that can happen, the chance that good idea is gonna get from the working level all the way through and be approved is very minimal. The limit on below threshold reprogramming definitely seems like it's out of date. I totally understand the power of the purse, and the limit of a 20% threshold seems like something that could work, but the $10 million limit just doesn't work.

The same interview subject also addressed how a lack of a management reserve means that the budget may not be an accurate reflection of what can be expected as a project evolves:

> We're not supposed to budget in management reserve. As the federal government, we're supposed to budget to target, not budget, to what we think is reality. So regardless of the fact that 0.5% of projects complete on cost schedule and with existing requirements, we're supposed to budget to that existing kind of spot versus being able to budget to that kind of management reserve. So, we're not necessarily allowed to put in those planned unknown rework steps that we know is gonna happen.

LUSV/MUSV Program Key Finding #2: A one-size-fits-all PPBE process does not work well for new technology programs with no significant cost or development history. Interview subjects spoke about how the PPBE tends to be monolithic and fails to adequately distinguish between major capability acquisitions and programs that need to adjust to rapidly evolving technologies:

> We reformed the acquisition system, but we didn't reform the associated budgeting system. There are things about the PBBE process that do work. I do think [it works] in major capability acquisition where you're buying very large, very slow-moving things like ships. It sets up a nice structure with nice guardrails that allow you to get a highly complex, very large amount of money committed … Now you're using this process that works great for buying billion-dollar things for things that cost hundreds of thousands of dollars or small millions. It's the one-size-fits-all process that quickly becomes onerous to the point of, almost, we work around the system rather than let the system work for us.

The interview subjects for LUSV/MUSV emphasized the difficulties of cost estimating for programs with evolving requirements or with new technologies having little to no precedence or budgeting history. In particular, they detailed the challenge of explaining the unique nature of new technology programs to other entities like the U.S. Government Accountability Office (GAO) and Congress. Interview subjects characterized those parties as relying on “napkin math” to judge their program primarily from budget books without appreciating the full story of how hard it is to plan and cost estimate for such a program.

Another example was provided by interview subjects about how the PPBE process can interfere with strategic planning for a new technology program. The Navy received a one-time congressional add of $42 million in 2019 to make an early purchase of a prototype four years earlier than serial production had been scheduled to begin. The funding was described as poorly aligned with the Navy’s strategy because, as a one-time add, the money was received too early to fit into the overall development timeline for the program. Ultimately, the single boat unit purchased by the Navy was considered an “orphan child” for which the acquisition documentation, long-term budget, cost estimates, and contracting pieces were ill-prepared to accommodate.

LUSV/MUSV Program Key Finding #3: J-books are not realistic for projects with many interrelated parts because they appear as an “à la carte” menu. According to the observations of personnel associated with LUSV/MUSV, the individually segregated budget lines of the J-books tend to convey a wrongful impression of how projects with multiple interrelated parts function. One Navy program office interview participant described how the J-books might appear to congressional staffers and result in congressional marks that impact program funding:

> The budget justification books appear like an à la carte menu, and that’s not reality. The budget books, the way that we’re supposed to break it down for staffers is: “here’s how much [I’ll fund] each of these individual items when actually those items are interrelated. And then they mark a portion of it that they think equates to that exact line item which actually breaks several other areas. … The marks are a huge problem, and the way that they mark it in that à la carte menu is not directly how we’re gonna be able to apply it.

Another government interview subject had observed congressional staffers failing to understand economies of scale when making budget cuts or appropriately pricing quantity units, assuming the same per-unit price to hold even when fewer quantities were purchased. He provided a hypothetical example, warning that he has seen this occur repeatedly in the past, whereby they were originally going to buy three units at $10 million each, but Congress only permitted purchasing two units. Moreover, contractor interview subjects also observed that congressional marks made during the evolution of the project appeared to reflect simplistic assumptions that did not fully appreciate the integrated nature of the hardware and software requirements.

Case Study #2: Air Force Collaborative Combat Aircraft (CCA)

The Air Force’s Collaborative Combat Aircraft (CCA) effort aims to develop unmanned combat air vehicles which are capable of operating as either tethered (in combination with) or untethered (operating autonomously) to manned combat aircraft. It was described in the FY 2024 Air Force Budget Justification Book as a program for “un-crewed weapons systems capable of enhancing crewed weapons systems to achieve air superiority.” The effort comprises part of the Next Generation Air Dominance (NGAD) initiative to develop advanced sixth-generation jet fighters. Per the FY 2024 Budget Request, CCA is intended to augment the advanced platforms through providing lower cost, complimentary systems to increase lethality in contested environments.

CCA’s unique creation was an administrative realignment rather than a budgetary new start. It is derived from the Autonomous Collaborative Platform program element (0207179F), which first appeared in the FY 2023 budget justification book as a continuation of previous work accomplished under the Skyborg Vanguard Program for integrating artificial intelligence into autonomous unmanned air vehicles and enabling teaming capabilities (Department of the Air Force, 2021). The core effort was supplemented by two ancillary programs in FY 2024 and now comprises three major lines of effort to develop and test an artificial control system dubbed the "autonomy package" (Harper, 2023). CCA began concept exploration, integration studies, technology risk reduction efforts, and prototyping in FY 2024, and the Air Force plans to spend more than $6 billion on CCA through FY 2028 (Harper, 2023).

CCA Key Finding #1: High levels of coordination with other government entities and commercial partners were integral to effective operations. Interview participants emphasized the importance of working closely with science and technology (S&T) partners from whom CCA's developmental technology was inherited, with the Navy on current platform interoperability concerns, and with Cost Assessment and Program Evaluation (CAPE) on budgeting concerns. In particular, extra effort had been needed to keep the multiple collaborating bodies cognizant of common standards for joint platform and software development.

Although the interoperability concerns were not driven by the budgeting process, PPBE appeared to have a segregating effect on the various agencies and add further layers of consideration to the strategic planning processes in inter-service efforts to develop quality joint capabilities. While interview participants expressed positive views regarding Air Force–Navy collaboration, they did notice that divergent acquisition strategies could lead to uneven budget line funding among different services if programs like CCA become a higher priority for the Air Force than the Navy, for instance. He noted that the process "introduces a lot of uncertain and risk if one of those budget lines doesn't get funded. It actually affects the overall outcomes for both services."

The CCA program benefited from the Air Force's close collaboration with industry. While the PPBE process tends to fix attention on winner-take-all efforts in the contracting realm and standalone defense projects at the funding level, the CCA comprises part of a high-tech, major joint capability solution as opposed to a single stack of components for one innovation. Thus, CCA benefits from what was described by one interview participant as "the momentum of all of industry going after this problem of fielding a capability, not just a single platform." The Air Force reinforces competition to build a pool of preferred vendors for various key technologies, and it anticipates a large contractor base of at least 35 companies for the program. To maintain executive control over the large industry base and harness the benefits of the extensive competition, CCA personnel maintain independent relationships with both the primes and their suppliers. Interview participants described the various ways in which the Air Force has sustained its active leadership while making efforts to involve the contractors in development, which requires it to keep them each informed of the CCA program's strict budgetary timelines.

CCA Key Finding #2: The PPBE process can interfere with service strategy. Several interview subjects associated with the CCA program expressed concerns that the PPBE process can occasionally be used by Congress as a tool to maintain control, causing project outcomes to deviate from Air Force strategy. Congress plays a powerful role in the PPBE process—it authorizes and appropriates the amount and timing of funding for various DoD activities in all phases, and it provides the limited authority for the DoD to transfer and reprogram funds (McGarry, 2022). One interview subject offered an example in which Congress, through its actions during the PPBE process, might have affected Air Force strategy and program process:

> The Air Force decided that it didn't want to spend a lot of money on 4th-generation capability development and backed away from the program. So, we offered a lot of money up in the omnibus to reallocate those funds somewhere else in the Air Force. Congress came back and denied the source because they want us to go fund 4th-generation and electronic warfare capability. They use the PPBE process to force us down a path where we don't think strategically about we should be going. And then it does have second, third-order effects on programs like CCA and NGAD because we're forced to try to figure out how we make it work at the portfolio level.

In spite of the potential problems posed by Congress' role in the PPBE process, interview participants acknowledged that the Air Force personnel were ultimately responsible for navigating the PPBE process to ensure program success. Moreover, regular communication with Congress can assist in helping to smooth over programmatic issues—such collaboration was identified as instrumental in the successful use of a technology transition process, rather than new start, to build the CCA program. Interview participants explained that a key to the success of a defense program such as CCA lay in maintaining a fine balance between congressional oversight and congressional overstepping.

CCA Key Finding #3: A flexible budget structure helps with navigating the PPBE process. Unsurprisingly, with a lack of flexibility often cited as one of the main issues with the PPBE process, CCA interview subjects spoke about the importance of budget structure in enabling greater flexibility. One reason for this was the improved ability to reprogram or move funds. Embedding CCA within the same program element of the larger NGAD initiative allowed money to be easily shifted between the different lines. On the flip side, one interview participant expressed concerns about the separate program elements belonging to the two ancillary programs of CCA, the Experimental Operations Unit (EOU) and the Viper Experimentation and Next-Gen Operations Model (VENOM). Due to the interrelated nature of the different lines of effort for CCA, they could not be separated from one another at the operational level even if the PPBE process could potentially cause them to be treated as isolated projects.

According to interview participants, a flexible budget structure facilitates rapidly evolving technologies but conflicts with the highly structured PPBE process. Cost estimating new technologies was one result, although the Air Force interview participants acknowledged the impossibility of a perfect budget estimate and the need to make progress without it, noting that "if you let everything shake out and try and get a perfect answer all the time, then you will never field a capability, and that's the only reason we have jobs."

CCA Key Finding #4: Program prioritization by leadership is a critical factor for successfully navigating potential budgeting or political-related issues. According to interview participants, CCA is unique from other Air Force programs and owes a large part of its success to how it has been driven by top-level leadership. Two different subject matter experts explained:

> CCA, is a little bit of a red herring in this conversation because, frankly, it's Frank Kendall's number one priority coming out of the Operational Imperatives. I think we were the only program that got fully funded as part of the process, so that made it a little easier … So, you give us the flexibility and you give us the access to leadership, and we can do things pretty quickly.

> It takes having priority access at the top level, and then the brute force at all levels—Air Force staff level, CAPE level, comptroller level, on the Hill—to be able to execute effectively within the PPBE process.

Case Study 3: Army Robotic Combat Vehicle (RCV)

The Army's Robotic Combat Vehicle (RCV) program, also referred to as the Remote Combat Vehicle program in the FY 2024 Budget Request Overview, is developing autonomous and semiautonomous ground combat vehicle prototypes, including the advanced autonomy and artificial intelligence algorithms to support them (GAO, 2020). The Army had originally planned to develop three RCV variants: Light, Medium, and Heavy (RCV-L, RCV-M, RCV-H). However, the Assistant Secretary of the Army for Acquisitions, Logistics, and Technology (ASA [ALT]) recently stated that the Army plans to focus on RCV-L development and will defer RCV-M development for the near future (Feickert, 2023). The RCV program is one of four signature efforts which are part of the Next Generation Combat Vehicles (NGCV) family of ground combat vehicles intended to prioritize rapid development and modernize the existing fleet.

As interview participants explained, the RCV project progressed in a non-linear fashion due to experimentation results, reaching BA 5 status for Development & Demonstration before regressing to BA 4 status for Advanced Component Development and Prototypes and returning to BA 5. With regard to its budget structure, RCV is fairly consolidated—it contains just three lines of effort, including an MTA Rapid Prototyping program, which are contained in one project comprising a single program element. The Army is using multiple competitively-awarded, consortium-based other transaction authority agreements (OTAs) awarded to various contractors to conduct experiments testing and building RCV technologies, and it is planning a down-select to a winning vendor to deliver prototypes.

RCV Key Finding #1: The PPBE process is not necessarily optimal for progress, but it is also not always a hurdle to operations or strategy. Key interview participants associated with RCV expressed observations that other factors had impacted strategic and funding decisions more than budgeting concerns. Major changes to the RCV program were enacted primarily due to non-PPBE-related concerns. For example, interview participants explained that experimentation outcomes played more of a role in major decisions, such as the core strategic decision to shift focus away from developing the light, medium, and heavy variants in order to prioritize the common light chassis to be adapted into the other variants later. One interviewee observed:

> PPBE did not impact the decision to focus on a common platform. It was the second phase, what we call the soldier operational experiment phase two, that was completed about a year ago and from that came the recommendation to shift the strategy for RCV … The feedback from the experimentation, that's probably the most significant piece, combined with it being an investment decision on maturing a capability before expanding on it. I think that is more what's driven it rather than, you know, we didn't have enough money, or we were concerned about being able to justify requests.

Major programmatic changes like the RCV-L prioritization were also attributed to program requirements, according to another interview participant:

> When you looked at it holistically and saw the gap between the light and the medium, it related back to requirements and not necessarily to budgeting at all. When the requirements community changed and we transitioned to a common platform, because we kept separate lines of effort in the totality of just an RCV single budget line, it allowed us flexible space to not delineate between an 'L' and an 'M' and instead focus on the common chassis-type platform, like in our recent public solicitation. So, it's not that we didn't want to look at it [RCV- M] or that PPBE hindered us from exploring it, it's just the investment at the time and the capabilities that go back to the requirements didn't really warrant it given where the Army wanted to go. Personally, I don't see an issue with [PPBE] at all.

One criticism regarding PPBE was that it might have slightly delayed prototyping, but not significantly so. According to one of the interview participants, the real challenge lay in communicating program plans and relating them to resourcing requirements rather than the PPBE process itself. The subject matter expert further explained that good lessons were occasionally learned through PPBE-related delays to obtain better information for informing experimentation and investment.

A different interview subject was harsher in his critique of the PPBE process, describing it as too “archaic” to keep up with emergent technologies and needs. Although he acknowledged it was a deliberate and structured approach, providing potentially important insights to the top-level leadership, he found it in general to be “out of touch with reality in terms of how funds are executed, how emerging needs present themselves, and how we adapt to them.” However, the interview participant expressed complimentary views regarding the PPBE process that were similar to those expressed by other subject matter experts, such as PPBE’s inadvertent effect of prompting higher-level strategizing about resources, which contributed to the Army working towards a common chassis rather than RCV-L and RCV-M variants.

RCV Key Finding #2: To facilitate programmatic success within the PPBE process, more frequent interactions with Congress are preferable. Key perspectives from RCV interview subjects endorsed more frequent interaction with Hill staffers to enable the program officers to provide more updated context for evolving program strategy, better explain any rapidly changing requirements, and educate the staffers on certain nuances of contracting to fill in any gaps of understanding. All three senior interview subjects expressed several concerns related to congressional relations, observing program offices’ inability to continuously update Congress, and Congress’ tendency to not inquire on program activities unless something was wrong.

As noted by the RCV subject matter experts, significant changes could occur in the program in the nine months between submission of budget exhibits and staff briefings and the budget being passed into law. As a result, acquisition professionals would sometimes only “get one bite at the apple every year” to deliver their message in March, without any meaningful Hill reengagements thereafter. With regard to the RCV program, this issue could be compounded by the fact that many Hill staffers seemed to lack sufficient training and understanding about the nuances of contracting, such as how critical acquisition authorities work in practice. This would result in budgets being marked for under-execution, particularly with activities using acquisition authorities that were critical to RCV’s agility and success. An interview subject provided the example of OTAs, which can go from zero to 100 obligated in a single day, but which lent themselves to perceptions on the Hill that the RCV program was not spending enough if the balance of zero remained too long.

As noted by the RCV subject matter experts, significant changes could occur in the program in the nine months between submission of budget exhibits and staff briefings (February or March) and the budget being passed into law (typically the following December). As a result, acquisition professionals would sometimes only “get one bite at the apple every year” to deliver their message in March, without any meaningful Hill reengagements thereafter. With regard to the RCV program, this issue could be compounded by the fact that many Hill staffers seemed to lack sufficient training and understanding about the nuances of contracting, such as how critical acquisition authorities work in practice. This would result in budgets being marked for under-execution, particularly with activities using acquisition authorities that were critical to RCV’s agility and success. An interview subject provided the example of OTAs, which can go from zero to 100 obligated in a single day, but which lent themselves to perceptions on the Hill that the RCV program was not spending enough if the balance of zero remained too long.

To accomplish effective and frequent communications with Congress, it was also imperative, according to three interview subjects, for the Army to present more of a united front when it comes to relaying certain messages, or to at least ensure internal leadership was kept adequately informed of program progress. One of the interview subjects explained that "if my own bosses don't know what we're doing, all the successes we've had, and all the work that we're putting into this and how we're moving the ball forward, then I'm certain Congress doesn't."

RCV Key Finding #3: Greater flexibility in the PPBE process would be more suited to addressing agile acquisitions, specifically when dealing with iterative requirements and different colors of money. Although the RCV program subject matter experts did not see the PPBE process as a major hindrance, as explained by the first case study key finding, they did feel it was a challenge to match the structured PPBE process to a program with evolving requirements, which required a different mindset. While the PPBE process is often better suited to linear technological development and "full-stack," completed technologies, RCV is progressing in terms of new iterations of technologies, which requires a different mindset in terms of development and long-term planning for new technologies, and being comfortable that "over these years being an 80% solution is good enough for the time now, knowing that eventually we'll get to 100%."

Per the experiences of those involved with the RCV program, increased flexibility in the PPBE process might entail providing program leadership with more budgetary authority, which could allow program leadership to be more open about program budget management reserves without fear another actor in the PPBE process could take it away. It also might entail developing a "colorless" type of money distinct from other DoD appropriations categories, which can be devoted to emergent technologies or innovative programs at the service-wide level. Reprogramming was viewed as an insufficient mechanism for redirecting funds to accommodate new technologies and address new threats.

RCV Key Finding #4: Consolidating program elements helps in achieving greater program flexibility. As described in the previous case study finding for the RCV program, the highly structured PPBE process can benefit from increased flexibility wherever possible. Therefore, when it comes to budgeting for a cutting-edge program, one of the ways to achieve that is through keeping the budgetary lines of effort within a single program element. This facilitates the ability to move funds as needed for a rapidly evolving program. All three interview participants involved with RCV supported the idea of consolidating activities into a single budget line to handle the risks of evolving requirements, and the RCV program budget has been intentionally designed to retain all its lines of effort within a single project and single PE.

The purposeful budgeting structure of RCV was reported to have had several benefits. Firstly, it enhanced RCV's ability to deal with congressional marks in situations when there is a "mark that is unspecified or there is a cut to the program line for no specified reason," because the structure provides the ability "to move and decide where to take the hit internally." Secondly, the structure allowed RCV to adapt to new experimentation outcomes and priorities, and to potentially move monies back and forth between the autonomous program and the platform as necessary due to the hardware and software being so integrated. The interview subjects stressed that funds that can be transferred or executed on an as-needed basis are especially helpful during the developmental phase of programs, while budget-related limitations are more understandable for the Procurement phase of a program.

Case Study 4: Space Development Agency (SDA)

The mission of the Space Development Agency (SDA; n.d.) is to "create and sustain lethal, resilient, threat-driven, and affordable military space capabilities that provide persistent, resilient, global, low-latency surveillance to deter or defeat adversaries." To accomplish this mission, the SDA is developing and fielding a Proliferated Warfighting Space Architecture (PWSA) consisting of multiple layers (or "tranches") of satellite constellations providing navigation, surveillance, deterrence, defense, communication, and various other functions to the joint U.S. warfighter. SDA intends to launch at least five tranches of commercially-procured satellites, each developed through two-year cycles using the MTA pathway for rapid prototyping and fielding and enabled by artificial intelligence. The rapid cycles are part of a "spiral model" that facilitate regular technology upgrades and contract competition.

The SDA was created in 2019 by the Office of the Under Secretary of Defense for Research and Engineering, and it was officially transferred to the United States Space Force (USSF) in 2022. As such, it follows a split authority, with the director of the SDA reporting to the Assistant Secretary of the Air Force for Space Acquisition and Integration for acquisition matters and the Chief of Space Operations for all other matters. The SDA shares one PE with Space Command, in addition to launch costs that are separately funded by Procurement in the Space Force's Systems Command budget per the National Security Space Launch (NSSL) central contract. However, the SDA has retained a unique level of autonomy and maintains three of its own separate RDT&E PEs. It also does not use the JCIDS process for validating requirements—interestingly, it tends to conduct programming before planning, as it budgets for tranches before specific requirements are developed through a Warfighter Council.

SDA Key Finding #1: SDA's use of the MTA pathway and the agile, iterative incorporation of commercial technologies are central to rapid product delivery. Feedback collected from SDA personnel stressed that the organization's agile and iterative acquisition model remains its most important asset for achieving success in terms of technology transition and schedule adherence:

> SDA's unique model relies on speed to achieve its mission and represents a departure from big, slow, expensive acquisition programs. Our model works because it doesn't rely on delivering the perfect solution, which tends to focus on capability over schedule and cost and instead choosing to provide 'good enough' capability to the warfighter at the speed of relevance.

One criticism of the PPBE is that it encodes divisions between research, production, and operations, thus stymying iterative or feedback-based development (Greenwalt and Patt, 2021). However, the SDA's spiral model for prototyping and fielding emphasizes iterative development, which facilitates the agency's ability to adapt to new technological advances, as well as continuously benefit from the knowledge that is gained after each satellite tranche is launched. The SDA is able to incorporate its spiral model and combat the PPBE's tendency to stymie iterative development through two principal enablers—a heavy use of the MTA pathway, plus a strategy of acquiring relatively cheap commercial technology.

Case study interviews and feedback expressed that SDA efforts to improve upon satellite tranche deployments have largely been guided by what was learned through the execution of previous tranches. MTAs were highlighted as essential for the SDA to rapidly improve satellite technologies through incorporating lessons learned. In pursuit of successful rapid deployment, a key interview participant felt that "the greatest near-term risk to our model is failing to use the Middle Tier of Acquisition pathway, or any pathway that enables speed," as opposed to an "obsolete acquisition model and strategies that are no longer adapted to the new threat environment, or that fail to provide timely, effective, and credible solutions."

Case study research indicated that SDA's iterative approach to satellite development and its contracting efforts in the commercial sector have supported affordable operations. When satellites can easily be replaced in the future, it can reduce non-recurring engineering costs and accelerate fielding. SDA also benefits from innovative commercial sector technological progress and incorporates lessons learned through its leveraging of small business programs like the Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) programs. SDA's public sharing of its technology roadmaps to help companies determine what will be needed, and when. The agency fosters "full and open competition for each layer of each tranche, as much as possible. Through that model, we hope to create a reliable and predictable marketplace that allows industry to invest, plan, and compete on a predictable timeline while also avoiding vendor lock."

Although SDA prioritizes commercial sector technologies, it is not only commercial technology that SDA adopts. It also provides a suitable environment for other cutting-edge projects that originated from the government. One example of this was when the Defense Advanced Research Projects Agency (DARPA) initiative FOO Fighter was transferred to SDA when Air Force partners were averse to taking on the technological risk of the new program. FOO Fighter failed to be included in the Space Force POM during two annual budget cycles, until Secretary of the Air Force Frank Kendall became alert to it and incorporated it into one of his operational imperatives. Facilitated by this critical endorsement of top-level leadership, SDA has been able to support maturation of FOO Fighter technology and is ensuring it will achieve funding:

> SDA is the ideal transition partner for mature capabilities … FOO Fighter was seen as a technology that could be incorporated into program architectures in the future. The cultural design of SDA is why FOO Fighter is with us right now.

Several challenges exist for the SDA to maintain its agility in the future, including possible new MTA reporting or other requirements levied on MTA use. One interview participant noted that the creation of an MTA advisory board has already slowed the timelines for MTA approval. SDA interview participants emphasized that while SDA's spiral model could certainly provide an example for other defense agencies to potentially follow, it is most likely not suited for large, exquisite systems, and conversely, SDA (plus other programs and organizations that rely upon the MTA) should avoid the bureaucratic red tape that could undermine agility and make SDA "a lot more like a major capability program."

According to interviews, SDA could potentially face another unusual challenge that contradicts the common experience of other DoD agencies. While many defense programs typically lag behind the commercial market in their incorporation of advanced technologies, the industry segments that SDA relies upon, such as those for optical communication link satellites, do not yet exist at scale to produce for the agency. Thus, SDA must nurture and build up its commercial supply chains, because "no matter what you do to the budget process, you can't acquire things on our timeline if industry is not prepared to respond."

SDA Key Finding #2: Due to SDA's mandate to rapidly deliver capabilities, budget requests must be made before requirements are finalized—programming occurs before planning. In interviews, SDA personnel stated a contrast between their organization's mandated delivery timelines and PPBE process timeline. An SDA interview subject explained that when the agency is acquiring capabilities at speed every two years, a one-year slip in funding in the Future Years Defense Program (FYDP) "can't be absorbed the way it is in legacy programs."

For the SDA, the PPBE's budgetary timeline is ill-suited to support emergent discoveries and new findings during program execution. To remain true to its core mission and maintain the two-year cadence, SDA technically conducts programming before planning. The requirement for a tranche is endorsed by its Warfighter Council six months prior to acquisition as opposed to 2.5 years. With a compressed schedule, SDA determines the budget for each tranche before actual requirements are known. SDA personnel felt that "reversing" the Ps of PPBE has worked well for the organization, but that its unique approaches have still posed unique challenges juggling cost estimates, requirements, and potential budget changes when budget planning for one tranche while working on another tranche of satellite capabilities and fielding a third.

SDA navigates its compressed timeline through detailed communications with Congress. This case study finding for SDA aligned with the Commission on PPBE Reform's interim and final reports, which emphasized effective engagement between SDA and Congress as an important cornerstone of efficacious transparency. Through frequent staffer engagements, SDA builds trust by providing detailed cost and work structure breakdowns, including comparisons between original cost estimates and actual cost outcomes—for each tranche, each performer on contract, each program element, and other project details.

SDA Key Finding #3: PE consolidation gives SDA more flexibility to successfully navigate program developments, but external stakeholders who seek to impact programs sometimes prefer a divided PE structure. Although the SDA was officially transferred to the Space Force in 2022 for administrative reasons decided by the Department of the Air Force, it has managed to retain the autonomy it has enjoyed since its inception. One of the most important aspects of this autonomy is the organization's budgetary structure of "large PEs encompassing multiple programs" which, according to SDA personnel, allow for the greatest flexibility to move funds when needed to ensure timely mission success—Section 1601 of the FY 2021 National Defense Authorization Act (NDAA) specifies that the SDA's PEs should remain separate from other Space Force programs.

SDA has opposed congressional and DoD efforts to split ground funding for its programs into separate PEs, which it has attempted in part as bids for increased control or transparency. SDA personnel explained that the Space Force attempted to create a separate PE for ground systems, in the same pot of money with Space Systems Command's (SSC's) Medium Earth Orbit-Ground, leading to "a big lump of money for SDA and SSC to figure out." Through increased staffer engagements with Congress to provide more detailed work and funding breakdowns within each of its PEs, plus referring to original statutory language to fight the loss of one of its distinct PEs, the SDA intends to achieve the realignments of funding to its primary program PEs. As one SDA interview participant noted, the ability to manage space platforms, ground stations, transport layer operations, and integration all in single or fewer portfolio-oriented or mission-based PEs allow the program office to be more responsive to events or proactive ahead of challenges.

SDA Key Finding #4: Building and launching SDA tranches can be challenging to manage in existing budgetary categories. SDA capabilities are rapidly developed with RDT&E monies, while the launch vehicles that deliver those capabilities are funded by procurement monies. SDA's management of different appropriation categories or "colors of money"—specifically procurement and RDT&E accounts—has been essential, and occasionally challenging, to the agency navigating PPBE processes while achieving the agency's mandate. Interview participants explained how the split funding impacted operations in fiscal years 2022 and 2023:

> Congress decided they wanted us to accelerate the fielding of our Tranche 1 track. They wanted us to speed up getting to a capability that could cover INDOPACOM by about a year. And they gave us a significant amount of money over the course of two years to do that. What they didn't do was fund the associated launches that go with that. And by the time appropriations passed, it was too late for us to then program for those launches because the budget was already headed to the Hill and closed out. The process for the next fiscal year in which we would have to acquire the launch to go with that was closed.

The interview participant further explained that when the split funding results in discrepancies between different accounts supporting the same mission, it prompts the DoD and Congress to have to work together to resolve the issue, either reprogram the funds or appropriate a plus-up, potentially delaying program executions. In this specific instance, Congress eventually appropriated a plus-up to provide additional funding for the satellite launch. The anecdote supports overall case study findings that the budgetary structures and timelines of the PPBE process can make it more difficult to accommodate immediate operational needs, especially on an as-needed basis by the combatant commands.

Case Study 5: Tactical Intelligence Targeting Access Node (TITAN)

The Tactical Intelligence Targeting Access Node (TITAN) program is the Army's effort to develop a next-generation intelligence ground system to improve upon the existing capabilities of legacy ground systems, which will likely be phased out after TITAN is fielded. TITAN will ingest and fuse massive amounts of incoming sensor data from the warfighting theater, and it will be the first intelligence ground station to use AI and machine learning to classify sensor feedback to turn it into real-time intelligence for the warfighter and deliver it via lethal and non-lethal networks (Army Program Executive Office - Intelligence, Electronic Warfare & Sensors [PEO - IEW&S], 2021). Per the Army's FY 2024 RDT&E budget estimates, TITAN's deep sensing capabilities will support automated target recognition, identification, geolocation, and other functions that enable immediate situational awareness as well as long-range precision targeting and firing.

TITAN emerged as a modernization activity for ground station capabilities within the Distributed Common Ground System-Army (DCGS-A) program, which contains the elements for a vast amount of Army intelligence capabilities at the tactical, operational, and strategic levels. By FY 2022, TITAN was fully initiated into the acquisition process and executed as a rapid prototyping effort using the MTA pathway. As a natural evolution of previous DCGS-A programs, TITAN draws on expertise and resources from various Army organizations and initiatives, particularly the technologies developed through Project 907, known as the Tactical Exploitation of National Capabilities (TENCAP; Hitchens, 2020). Current funding for TITAN prototyping focuses on the advanced variant which will be used for heavier platforms such as tactical trucks and eventually be adapted to a basic variant designed for lighter platforms.

TITAN Program Key Finding #1: The use of a MOSA approach, the MTA pathway, and OTA contracts have led to rapid prototyping and program success in TITAN but still pose unique challenges. TITAN interview participants commented on the inefficiency of a two-and-a-half-year time frame from program offices' budget submissions to service headquarters before funds are made available to the program offices for obligation. However, TITAN's use of rapid and flexible acquisition authorities has insulated it from many negative impacts of extended timelines inherent to PPBE, JCIDS, and FAR processes. One government interview participant concluded: "there are a lot of tools in the toolbox, a lot of flexibility, and it's really on acquisition professionals to determine how to use those tools to best achieve what they are asked to deliver."

For a program like TITAN that heavily incorporates both hardware and software, PPBE's emphasis on planning is irrelevant when technological unknowns outpace the budget as it passes through multiple entities for adjudication, integration, review, and debate before obligation of funds. One interview participant expressed concerns that during this process, technologies can "go obsolete, or you know they're going to go obsolete in two years and you still want to buy, you have to buy it now because it is being programmed. But the replacement can be considerably more money that what you were planning for."

TITAN is heavily dependent on software technologies, and it is incorporating the use of a Modular Open Systems Architecture (MOSA) approach to iteratively incorporate evolving technologies. While the PPBE process can complicate a program's ability to navigate technology transition, acquisition approaches like MOSA can mitigate the hindrance of long timelines and guessing games for program needs. MOSA allows for system-compatible components to be iteratively added, removed, or replaced through the platform lifecycle, making it easier to keep up with the pace of technological advancements, incorporate ongoing soldier feedback, and avoid lock-in with proprietary solutions. Although MOSA can pose new challenges with regard to cost estimating for the plug and play of emergent technologies, it has been enthusiastically embraced by TITAN personnel as a means of working with, rather than against, uncertainty.

TITAN personnel also highlighted the MTA pathway as a great enabler of the program's speed, facilitating a rapid succession of prototypes and a maturation of relevant designs before final requirements documents are written. A TITAN program lead contrasted TITAN's use of the MTA pathway with the complicated timelines of typical major capability acquisitions:

> Say for example, if we had approached TITAN as a major capability acquisition program, we would've gone to a milestone B, we would've had an ADM [Army Design Methodology], we would've had an APB [Acquisition program baseline], and that would've established specific parameters for the program that by the time we initiated, we probably would've been oriented on that procurement funding and a lower RDT&E number than where we've gone to in the rapid prototyping program. Which would've led to us initiating a program and then probably doing a significant deviation or a breach within the first year, because we learned so much in the first six months that caused us to have to look at different funding alignments and what we would resource and program. And that's where the interaction of PPBE and the MTA approach was beneficial to the TITAN program, otherwise we probably would've had to re-baseline program at least once, already, in the first 15 months of this program.

A major aspect of TITAN's system design and contracting approaches has been its use of down-select competition for prototyping, and it has awarded two major OTA contracts as part of its extended competitive prototyping effort by Palantir and Raytheon (Gill, 2023). Perspectives on OTA procurement authorities were nuanced in industry interviews. Understandably, the nontraditional contractor perspective placed greater value on the OTA's ability to level the playing field for industry competitors, while the traditional contractor perspective valued the protection afforded by structures of FAR-based contracts as opposed to OTAs. The nontraditional industry interview perspective designated the treatment of the OTA contract like a FAR-based contract as the biggest challenge currently faced from an industry perspective. The insertion of more FAR clauses was said to have reduced flexibility, increased bureaucratization, slowed funding timelines in the POM cycle, and hamstrung the government from moving forward more quickly with decisions and future phases due to fear of bid protest.

The traditional contractor perspective acknowledged the potentially negative impact of treating OTAs like FAR-based contracts but cited several issues inherent to OTAs. The first issue was the creation of middle barriers between government and industry, possibly leading to communication delays and loss of information in translation, as well as reducing direct collaboration. A second concern was that certain development programs might not be as effective because of their competitive nature cost driving relationships and adding an artificial element to contracts which do not always increase the pace of technological progress. Lastly, OTAs can necessitate a large cost-sharing component, impacting a traditional contractors' ability to innovate.

TITAN Program Key Finding #2: TITAN has benefited programmatically and technologically as a continuation of previous Army research efforts and funding lines. TITAN personnel explained that the program has faced fewer challenges in the PPBE process in part due to its privilege as a high-priority, high-profile program derived from a major legacy program. As such, it has enjoyed robust Army and other government support which has facilitated coordinated progress, and it inherited a more streamlined budget structure which contributes to program agility and success. One interview participant believed that if the program was a new start, the Army would still be waiting to start advancing the program and obtain dedicated funding lines.

> If we had completely approached this from a traditional method, we probably would just be barely starting TITAN in ’24 if we had truly followed the full PPBE process for initiating. But we were able to find ways in 2020, 2021 in particular, to begin the program and start doing work to advance it in advance of having funding lines specifically for TITAN. And that was done through coordination with OSD, with Congress, and everyone else being very transparent on it. We would begin applying money within the scope of existing programs towards these future requirements.

Thanks to the advancements of its parent programs plus strong ties with other Army entities, TITAN has avoided the hassle faced by new programs waiting for their turn to be rolled into a crowded portfolio of existing programs of record during a service’s POM process. It has also leveraged the technological expertise of other programs and organizations to save money and move more quickly without growing an entirely new workforce. Interview participants were keen to highlight BA-4 as an effective transition vehicle for segueing technologies from the TENCAP 907 line into TITAN, especially for the space-based component of TITAN’s ISR function. One interviewee described the TITAN-TENCAP link as a “habitual relationship where they’re an incubator as new space technology comes online. There’s a logical bridge there over the valley of death where it’s a natural transition from the TENCAP Office into the TITAN program of record.”

Interview participants further noted that while TITAN benefits from the increased agility of a more streamlined budget structure than its parent DCGS-A program, it can still be challenging deciding how to present funding for congressional justifications. One interview participant described the balancing act of providing Congress with necessary budget information:

> For a program like TITAN, having one funding line is helpful. With DCGS, which was twice the order of magnitude (it was a giant program, ACAT I), that can look like just a large bill fare for the rest of the Army when you have one giant funding line. And now they will take that to pay bills and the PMs are left to figure out how you execute the remainder of the dollars. Having more specificity is nice because it makes it easier to defend cuts to one of our individual program lines, but it also kind of locks you in; you just have less flexibility. So, it’s definitely a balancing act in how we write our P and R Forms and how we lay out funding lines. It’s a little bit of an art and a science.

TITAN Program Key Finding #3: The shift of program funding from Procurement to RDT&E, accomplished with effective stakeholder alignment, ensured that appropriate investments were made in prototyping but had downstream effects on industry efforts. Early on, the need for Procurement funding for FY 2024 had been overestimated, and as TITAN evolved, the Program Office recognized that RDT&E funding was more appropriate to developing and integrating new technologies into the program. Annual President’s Budget requests for the TITAN program were changed between FY 2023 and FY 2024. While FY 2023 projected $298.9

million in procurement funding for FY 2024, the actual FY 2024 budget materials included zero procurement funding.

The consensus among the interview participants was that the funding realignment was done quickly and without any negative impact on the TITAN program – largely due to effective stakeholder engagement. The staff noted this success as an example of Program Office “collaboration across the enterprise, between Army and OSD and the Hill to work on right-sizing the funding lines … it was the Army speaking with a unified voice, it wasn’t a bunch of different opinions … It was quite effectively communicated and supported at all levels.” The interview participants placed particular emphasis on an engaged and positive relationship with Congress, marked by regular communications, as an enabler of flexible execution. When funds needed to be moved from Procurement to RDT&E during the budget phase following the President’s budget submission, Congress was asked to change the budget, marking budget lines and moving money, before the budget submission was approved prior to the year of execution. Funding adjustments were made within the portfolio without adding dollars to the program.

Both the literature review and the interviews highlighted that program managers are graded on how they spend appropriations and are penalized (i.e., “dinged”) if they don’t spend within a prior ordained timeline. As a consequence, the speed of spending is often a greater concern than the effectiveness of the spending. As a result, spending decisions are made which could easily be called into question if the standard was effectiveness, efficiency, or even performance, rather than whether money is spent according to a quarterly sequence. Often, as exemplified by the TITAN Program Office’s decision to extend the prototyping phase, it is more important for the PPBE process to allow for flexibility to change course when needed, rather than for finding ways within PPBE to accelerate technology transition within the program of record. One interview participant contrasted TITAN with other Army programs as such:

> Because we’re now in a different situation, we’re doing a lot of prototyping and we’re using a lot of RDT&E to buy that hardware. And the problem with that is you do not have disbursements in your RDT&E until you receive that hardware. I don’t think that’s an issue per se with the TITAN program, but we have other programs where it might take two or three years to receive a piece of hardware, and if you don’t have disbursements showing that as you’re going through your under-execution with OSD, you’re getting dinged constantly.

Although the funding realignment was accomplished with relative ease and without noticeable impact to TITAN’s early success, industry interviews captured some of its downstream effects and potential implications. The realignment, due to initial overestimation of Procurement funding needed for TITAN at the start of the PPBE process, adds an additional 18–20 months to the competitive down-select process. It is estimated that approximately one-third of the delay, comprising the final months, will be caused by government deliberation on selecting the winning contractor. While the extension of the competitive cycle and additional months required for decision-making could support TITAN’s development efforts, the resulting timeline delays have several implications.

One industry interview participant highlighted the final months of government deliberation as the main culprit for any potential negative impacts of a delayed technology transition. He explained that “not only is it delaying program progress, but [the government is] also spending extra money to keep both vendors on an additional 18 months before they can actually move to an award decision.” Another interview participant working for a different contractor rationalized that such delays are not unique to the TITAN program—rather, that they are inherent to development efforts that occur alongside competitive scenarios. It occurs due to a variety of reasons, such as difficulties balancing fair competition between suppliers with specific standards for innovation and broad requirements.

Case Study 6: Joint Rapid Acquisition Cell (JRAC)

The Joint Rapid Acquisition Cell (JRAC) is a small organization within the DoD that is uniquely positioned to coordinate with the services (and sometimes DoD agencies) in helping them fulfill their mandated obligation to fund, deploy, and sustain solutions for the urgent operational needs (UONs) of the warfighter within a rapid time frame. UONs, defined as capability requirements impacting contingency operations, originate from combatant commands and are further classified as either Joint Urgent Operational Needs (JUONs) for ongoing contingency operations or Joint Emergent Operational Needs (JEONs) for anticipated contingency operations (CJCS, 2018). As JUONs/JEONs must be reviewed and validated at multiple levels of authority, represent combatant command priorities, and require the services to make tradeoffs in their defense portfolios, the threshold for their approval is high: the potential for unacceptable loss of life and/or critical mission failure if the capability is not provided. The mission of the JRAC is twofold: firstly, it should facilitate the resolution of JUONs or JEONs through the designation of the DoD entity (almost always a military service branch) responsible for funding and filling the operational capability gaps. Secondly, the JRAC must monitor and ensure the timely fulfillment of the solution from development to sustainment, helping to resolve issues that arise as the UONs transition into a program of record.

Comprising one of several measures designed to meet the demands of asymmetric warfare during the wars in Iraq and Afghanistan, the basic structure of the JRAC was established in a 2004 memorandum emanating from the Office of the Deputy Secretary of Defense (Middleton, 2006). Although the JRAC itself cannot fund a capability or roll it into a program of record, it is equipped to accommodate the urgency of operational needs through a set of acquisition and funding authorities known collectively as Rapid Acquisition Authority (RAA). The JRAC uses its RAA to help translate the operational priorities of the combatant commands into the POM cycles of the services, and it allows certain DoD components to make use of available funds on a flexible basis without following the typical phases of the planning, programming, and budgeting phases of the PPBE (GAO, 2023). Along with its RAA, the JRAC is also empowered to facilitate rapid acquisitions for JUONs and JEONs by serving as a single point of contact and intermediary for critical decisionmakers within the DoD.

JRAC Key Finding #1: JRAC efforts highlight the challenges of developing and deploying urgently needed capabilities to support operational needs via the services' respective PPBE processes. A critical takeaway from interviews was that even after the JRAC has handled the initial difficulties of coordinating with DoD leadership to validate UONs and designate a service (sometimes a DoD agency) for incorporating a capability, there are often many delays and difficulties adapting the JUONs or JEONs into capabilities within the services' portfolio. These challenges adapting urgent operational needs stem from the difficulty of capturing reliable new funding lines through PPBE and the nature of PPBE which induces the services to segment their own priorities separately from the Defense Department and combatant commands.

Despite the JRAC's unique authorities and position to assist with the fulfillment and funding of JUONs and JEONs post-validation, the JRAC does not duplicate the functions of service-unique rapid acquisition processes, and once a service adapts a JUON or JEON into a capability within its portfolio and obtains the relevant new funding lines for it, the schedule to deployment is contingent upon the service. Moreover, despite the JRAC's RAA and the statutory requirement for DoD components to address JUONs or JEONs, the JRAC lacks a forcing mechanism for the services to turn a possible solution into a program of record.

A recurrent theme throughout interviews was that the services are averse to adjusting their programming and budgeting for UONs because it requires them to make tradeoffs affecting their priorities. The interviews suggested that funding for ongoing military service modernization efforts are among the most selected sources of quantity and funding cuts to accommodate the solutions. Since services designated for fulfilling a JUON/JEON are responsible for full lifecycle costs, the JRAC sometimes struggles to insure against "drive-by acquisition," a phrase that was coined to describe instances where material solutions are fielded to the warfighter without adequate Service support for long-term program management and oversight (Middleton, 2006).

An example of drive-by acquisition in one recent scenario in which the JRAC helped oversee the development of a hardware solution, valued at approximately $25.7 million, that was initiated in FY 2018 and sustained through FY 2021. Although the Service designated for oversight had identified the system as a future program of record to be fielded until FY 2025, it chose not to fund further sustainment after the capability lost its JUON designation. To fund continued sustainment, the Service would have had to reduce the PE funding levels and quantity purchased for another unrelated system, which it was unwilling to do. The Service explained that it was "focusing more on strategic long-term modernization priorities, not a short-term band-aid solution." Thus, sustainment funding could not be captured through the PPBE process.

The UONs' budget battles with preexisting service programs of record are a byproduct of PPBE shortcomings because the long timelines of the PPBE process inherently make tradeoffs harder. The PPBE process' emphasis on maintaining or adding funds to prior programs of record negatively impacts the services and other DoD entities attempting to deliver important capabilities to the warfighter within a timely manner. One non-JRAC interview participant with both industry and government experience had observed that in general, programs were rarely cut or slowed down, even when not executing well, "in hopes that these programs would deliver something, someday." Additionally, the interview participant observed an ongoing scenario of a government organization "waiting out the PPBE cycle" to incorporate a new program architecture because ongoing programs of record had left no room in the organization's budget.

Another part of the difficulty transitioning JUONs and JEONs into military service programs of record might be attributed to a disconnect between the combatant commands and the services. Across multiple case study research efforts, interview feedback from the JRAC and non-JRAC personnel associated with other defense programs suggested that the PPBE process for incorporating new funding lines can engender rigidity and lack of fluidity between combatant commands and the services. One JRAC interviewee described a potential dynamic that could occur when the commands' high-level aims are not in alignment with current service objectives:

> For something like long-distance ISR, you might hear from combatant commands that they want ISR [Intelligence, Surveillance, & Reconnaissance] that can remain in the air for days collecting data. The Air Force has the mission for the ISR, but can say they don't need it because they have the ISR program for the MQ-9 and want to focus on manned aircraft rather than unmanned. Meanwhile, the Combatant Command will tell you they need to expand the unmanned need. It then becomes a question of who becomes responsible in that force … Who's going to pay the new bill? At the end of the day what is the requirement? How many hours or platforms? How many people need to get it on this? … The services look at this in terms of having a new bill.

Delays and difficulties rolling urgent capability needs into the services' POM cycles can result in several negative repercussions. Since UONs are validated as high-stakes operational needs for the warfighter, their lack of fulfillment or sustainment, at worst, could negatively impact the mission or the safety of the warfighter. Moreover, drive-by acquisition can also waste money. If a program is not supported to the point of full operational capacity and sustainment, then the funds that went into its development or initial fielding were spent on a solution that had not been fully capitalized to the intended extent. It can also affect joint operations—even though JUONs and JEONs may be ultimately programmed into the budget of a sole military service, they are considered joint in that their importance overlaps service-specific missions (Middleton, 2006). Finally, the JRAC's difficulties transitioning solutions into viable programs of record can affect the incorporation of cutting-edge technologies, a recurring challenge across the DoD.

JRAC Key Finding #2: Phasing out OCO funding has made it increasingly difficult to secure funding to fill urgent capability gaps, especially JUONs and JEONs. The ability of the JRAC to fulfill its mandate has met with increased challenges since its original conception during the previous wars in the Middle East, when discontinued supplemental funding lines such as the Iraqi Freedom Fund often provided the primary source of funding for JRAC-enabled solutions (Buhrkuhl, 2006). In the past, the JRAC also utilized separate Overseas Contingency Operations (OCO) funding lines, which supported direct war and enduring operations costs and which were unencumbered by discretionary spending caps. JRAC came to rely on the use of OCO, which has shifted into base budgets, to subsidize development and sustainment of new urgently-needed capabilities. Today, JUONs and JEONs are almost exclusively funded by the services.

According to one interview with JRAC personnel, the average dollar amount for a JRAC-facilitated program or capability falls between $25 and $75 million—"small bites, by DoD standards." Despite the relatively small dollar value for most JUONs or JEONs, JRAC-facilitated solutions appear to face increased difficulties obtaining funding through the services' PPBE processes. These challenges suggest that the decline of OCO funding has impacted the ability of the DoD to address and sustain the immediate needs of the warfighter. One JRAC interview participant contrasted OCO-era funding with the current challenges faced by the services as they unwillingly balance their budgetary priorities with those of combatant commands:

> When there was OCO funding for both the Iraq and Afghanistan wars, sometimes that OCO would make it a lot easier to the services to where they didn't have to take funding out of their topline budget. In these cases, even sustainment funding didn't have to come from the services, it was all OCO. So that was certainly much less painful for the services that way because they didn't have to look internally and say they had to kill this ground vehicle program to fund this because this is de facto the highest priority of the Department, even though it doesn't align with our long-term modernization efforts.

When services or DoD agencies must draw from their own funds rather than a specialized funding line for critical wartime operations, the PPBE-related timeline lags and hurdles that hamstring regular acquisitions also become an issue for rapid acquisition processes, like those needed for the fulfillment of JUONs and JEONs within a reasonable time frame. As a result of the decline of OCO, the U.S. warfighter faces budgetary delays that cause it to be without critical capabilities for longer than intended.

JRAC personnel described a recent scenario, well past the era of OCO, in which funding difficulties caused major delays before one capability could be turned into a dedicated program of record and subsequently executed. In FY 2019, the JRAC helped to develop a JUON solution, at an estimated $28 million in cost, which was an adaptation of an existing software merging data from multiple sources. In the POM cycle, the software adaptation failed to find funding for two years, during which the warfighter had to function without this critical capability. The software was eventually rolled into and sustained through a larger Defense Information

Systems Agency (DISA) program. The solution eventually succeeded thanks to strong advocacy, proactive action by the JRAC, and the perceived importance of the capability:

> The system itself was very successful, and so it grew into a very useful program that the user community could rely on. And so there was a loud outcry of, “we can’t lose this capability,” and that was briefed back to us in program review. And my director, to the three-star that was responsible for funding it, told him to cut another program going forward, and that there was value in [the software]. And the three-star saw the value, understood it. It required person-to-person advocacy, but it was not a hard sell, it just required seeing it across the finish line.

As an addendum to this case study finding, it is important to note that JUON and JEON solutions still usually manage to obtain funding on time frames much shorter than acquisitions not facilitated by the JRAC. Per interviews with JRAC personnel, the average length of time from JUON/JEON approval to the obtainment of a funding line is still only two to three months. Moreover, there are other factors that distinguish how rapidly JUONs or JEONs are fulfilled, and difficulties obtaining funding also occurred at the peak of OCO spending, when a wartime sense of urgency prevailed. The impact of the decline of OCO should be examined within the context of the overall mindset shift that has occurred since the ending of conflicts in Iraq in Afghanistan, which has also most likely changed the services’ willingness to rapidly fund immediate warfighter needs through their respective PPBE processes.

Additional Conclusions and Recommendations

For all six case studies, interview subjects observed that the PPBE process had tangible impacts on technology transition and program success, but the extent and nature of these impacts varied substantially depending on the unique contexts of each DoD organization or program. Interestingly, the PPBE process did not seem to have as much of a negative or slowing effect on technology transition as the research team had initially hypothesized. Many program disruptions discussed in the cases were linked to the unpredictable nature of technological experimentation or other features of the defense acquisition system, like the requirements process. This suggests that in some instances, the PPBE process can indeed be a faulty scapegoat for the infamous valley of death. Nevertheless, for the technology-heavy programs and efforts discussed in this paper, PPBE was more likely to be viewed as an obstacle rather than an enabler of rapid development and deployment of new capabilities, and it could benefit from targeted reforms.

Many of the top-level findings summarized in the introduction to this paper were consistent throughout each of the six case studies. Strong senior leadership played a vital accelerating role for new or necessary technologies for every instance in which the PPBE’s perceived obstacles fostered the semblance of a valley of death. Interview participants associated with nearly every case study also expressed strong preferences for increased budgetary flexibility, enabled in part by budget structure and J-book organizational features—typically more consolidated, mission-focused PEs, and sometimes less division between appropriations accounts. The case studies were also generally characterized by a strong appreciation for agile approaches like the MTA, plus an emphasis on the importance of thorough, positive congressional engagement.

Interview participants generally perceived the PPBE as an annoyance to be dealt with or tamed however possible. Per the interviews, a need for increased coordination with combatant commands and government agencies were discussed as additional points of concern in the PPBE process with regard to technology transition, although these points appeared to be less immediately impactful than the need for regular congressional engagement. Interview participants, particularly those associated with CCA and TITAN, advised increased coordination with government agencies and adjacent program offices. Some also suggested increasing the authority of the combatant commands, including direct allocations of funds to the commands.

The budgetary inflexibilities caused by certain constraining aspects of the PPBE were strong contributors to negative views of PPBE among the government and industry personnel. To combat such constraints, several interview participants, particularly among SDA and LUSV/MUSV, advised adjusting or streamlining reprogramming thresholds to accelerate or accommodate the changing circumstances of acquisitions or capability delivery. They suggested increasing the reprogramming threshold limit or allowing for further reprogramming changes during the year of execution, after a program budget request for the next fiscal year has been finalized. Per the experiences of those involved with the RCV program, increased flexibility in the PPBE process might entail providing program leadership with more budgetary authority, allowing program offices to be more open about program budget management reserves without fear that another actor in the PPBE process could take it away. It also might entail developing a “colorless” type of monies distinct from other DoD appropriations categories, which could be devoted to emergent technologies or innovative programs at the service-wide or joint levels.

Along with budget structure, interview participants fully endorsed agile approaches like the MTA, OTA contract vehicle, MOSA, and other solutions, although successful use of these approaches is further impacted by other aspects of the acquisition process such as JCIDs, or industry efforts in the contracting realm. However, as illustrated by the JRAC since the decline of funding accounts like OCO, none of the procedural expedients to the acquisition and deployment of new capabilities can compensate for the necessity of a source of reliable funding. If service and congressional priorities do not support that funding, platforms will face a difficult road in technology transition, from development to deployment to maintenance.

Interview participants, both government and industry, observed that poorly performing programs often continued to receive funding at the expense of new capabilities needed, largely due to a reluctance to cut old programs. One solution to this issue might be to assess crowded legacy programs experiencing sprawl with their many funding lines and to utilize BAs to transform some of these funding lines into modernizing efforts that better align with the long-term priorities outlined in current service strategy documents. Per the case study interview perspectives, obtaining reliable funding and achieving service-wide and congressional support appeared to be a smoother process when new defense efforts were descendants of parent or legacy programs. TITAN provided an excellent model for such a transition as a well-supported child of the Army’s DCGS-A program through which it inherited legacy technological progress.

For many key enablers of flexibility, rapidity, and success in technological development and deployment among the six case studies, a large number were unexpectedly more likely to be hamstrung by restraints and bureaucratic tape which were not caused by PPBE-related restrictions. While adjusting the timelines of PPBE and increasing communication with Congress could facilitate technology transition, PPBE-targeted reform alone cannot alter the changeable nature of Congress itself as elected officials perform their legislatively endowed gatekeeping role. Nor can it prevent the unpalatable but unavoidable necessity in resource allocation and discretionary defense spending: difficult tradeoffs. The fine balance of navigating the classic military trifecta of modernization, force structure, and readiness has been the challenge for every society’s armed forces since the first soldiers were deployed to provide for the common defense.

Case study interview participants touched on a wide variety of non-PPBE-related and exogenous factors that should be taken into consideration to improve the success rate of capabilities traversing the valley of death. All case study sources, especially JRAC personnel, stressed the human element to finding solutions; the importance of advocacy during the POM process to advocate for operational needs, better accessibility to the relevant chain of command, and increased training. Interview participants conveyed experiences in which other key program participants displayed a lack of awareness of PPBE protocol, J-book structure, or other information that was simply critical to DoD functions—for instance, some DoD personnel were not even aware that they were required by statute and directive to undertake JUONs and JEONs.

In conducting the case study research, it was occasionally difficult to determine whether the so-called valley of death hampered key technological progress or whether it could have provided a useful screening process—not all prototypes or commercial technologies are viable for transition to operational use. The quality of technologies that are transitioned, rather than the quantity and speed at which they are deployed, arguably represents the most vital consideration in the DoD's resource allocation. There were occasionally positive implications to be drawn regarding the timelines and barriers of the PPBE process, potentially fostering more effective, pragmatic funding patterns. Depending on the perspective, positive implications may be drawn from FOO Fighter, as discussed in the SDA vignette. Once FOO Fighter was adopted by the SDA, progress was rapid, and SDA moved quickly to issue its first solicitation for the experimental space-based sensor technology derived from the program. FOO Fighter's lingering in the valley of death could be interpreted as a failure of PPBE, or it could be interpreted as a positive outcome in which the delay allowed for a successful new capability to find an alternative home within an agency that was more culturally compatible with its nature and intended purpose.

The PPBE Reform Commission's final report, with its prescriptions for a new Defense Resourcing System, suitably capture the need to retain elements of the PPBE that serve a critical purpose, while reforming its most glaring flaws, including the need to streamline programming and budgeting functions which overlap. The Commission's final report, in addition to findings from the six case studies, seem to convey that a complete overhaul of PPBE functions may not be necessary. However, they do endorse certain radical changes, and it is worth noting that programs like TITAN and organizations like SDA have found success by following radically different processes from typical capability acquisitions which hardly resemble a typical understanding of PPBE's limitations. Research findings (including further insights derived from interview discussions that were beyond the scope of inclusion in this paper) suggest that the SDA in particular could provide an effective model for other agencies with unique mandates to deliver technologies on a rapid timeline. Its ability to implement iterative improvements and its low-cost incorporation of commercial technologies enable it to circumvent the worst of the valley of death. Employing satellite technologies which are close to or already at full viability, it can escape many of the risk aversion tendencies and fears of elected officials and avoid extended prototyping. Nevertheless, as noted by the interview participants, rapid acquisition models are not necessarily suited to major capability acquisitions for large, exquisite systems.

If any further summary recommendation is to be derived from the case study findings, it is that in a peacetime setting, with new threats on the horizon from technologically advanced U.S. adversaries, stakeholders in the budgetary process should currently prioritize modernization efforts. This entails making space for cutting-edge programs rather than renewing POM cycles for legacy programs with underwhelming track records of performance. It also entails making space for the use of newer defense pathways or contracting strategies (i.e., the Middle Tier of Acquisition, or commercial acquisitions) while avoiding overregulating and applying new constraints which prevent these methods from functioning as they were intended.

These mindset changes will facilitate a more suitable pace of technological development and adoption.

Findings and recommendations derived from the six case studies should be taken into consideration with the understanding that the majority of programs and organizations discussed were relatively new and in the earlier stages of technology transition. As such, the full extent of PPBE's impacts on the defense programs discussed in this paper, as they progress to full operational capability or maturity, is not yet fully known. As such, this paper recommends further relevant research. Continued interviews and data-based analysis efforts are suggested in order to better isolate and assess the impacts of the PPBE process on these as well as other, mature defense programs that have demonstrated long-term successes or failures.

References

Army Program Executive Office - Intelligence, Electronic Warfare & Sensors. (2021, September 27). *TITAN brings together systems for next generation intelligence capabilities*. https://peoiews.army.mil/2021/09/27/titan-brings-together-systems-for-next-generation-intelligence-capabilities/

Buhrkuhl, R. L. (2006, December). When the warfighter needs it now. *Defense AT&L*, *35*(6), 28–31.

CJCS. (2018). *Charter of the Joint Requirements Oversight Council (JROC) and implementation of the Joint Capabilities Integration and Development System (JCIDS)*. https://www.acq.osd.mil/asda/jrac/docs/CJCS-Instruction-5123.01H.pdf

Commission on Planning, Programming, Budgeting, and Execution Reform. (n.d.). *Defense resourcing for the future: Final report*. https://ppbereform.senate.gov/wp-content/uploads/2024/03/Commission-on-PPBE-Reform_Full-Report_6-March-2024_FINAL.pdf

Department of the Air Force. (2021). *Department of Defense fiscal year (FY) 2022 budget estimates, Air Force justification book volume 1 of 3: Research, development, test & evaluation—Vol-1*. https://www.saffm.hq.af.mil/Portals/84/documents/FY22/RDTE_/FY22_PB_RDTE_Vol-I.pdf?ver=DGijGVofWq4jnTnOLuU5Bg%3d%3d

Department of the Air Force. (2023). *Department of Defense fiscal year (FY) 2024 budget estimates, Air Force justification book volume 2 of 4: Research, development, test & evaluation*. https://www.saffm.hq.af.mil/Portals/84/documents/FY24/Research%20and%20Development%20Test%20and%20 Evaluation/FY24%20Air%20Force%20Research%20and%20Development%20Test%20and%20Evaluation%20Vol%2 0II.pdf?ver=pYOQLrjX71gVe8w6FCJOwg%3d%3d

Department of the Navy. (2023). *Department of Defense fiscal year (FY) 2024 budget estimates, Navy justification book volume 2 of 5: Research, development, test, & evaluation*. https://www.secnav.navy.mil/fmc/fmb/Documents/24pres/RDTEN_BA4_Book.pdf

Eckstein, M. (2022, May 19). What's new in Navy and Marine Corps unmanned boats. *Defense News*. https://www.defensenews.com/naval/2022/05/19/whats-new-in-navy-and-marine-corps-unmanned-boats/

Feickert, A. (2023). *The Army's robotic combat vehicle (RCV) program* (IF11876). Congressional Research Service. https://crsreports.congress.gov/product/pdf/IF/IF11876/9

GAO. (n.d.). *Next generation combat vehicles: As Army prioritizes rapid development, more attention needed to provide insight on cost estimates and systems engineering risks* (GAO-20-579). https://www.gao.gov/assets/gao-20-579.pdf

GAO. (2023). *DOD benefited from financial flexibilities but could do more to maximize their use* (GAO-23-105822). https://www.gao.gov/assets/gao-23-105822.pdf

Gill, J. (2023, April 26). Army to test TITAN prototypes this summer as it moves toward downselect. *Breaking Defense*. https://breakingdefense.com/2023/04/army-to-test-titan-prototypes-this-summer-as-it-moves-toward-downselect/

Greenwalt, W., & Patt, D. (2021). *Competing in time: Ensuring capability advantage and mission success through adaptable resource allocation*. Hudson Institute. https://www.aei.org/wp-content/uploads/2021/02/Greenwalt_Competing-in-Time.pdf?x91208

Harper, J. (2023a, March 8). Air Force requests approximately $500M in fiscal 2024 to launch new CCA drone program. *DefenseScoop*. https://www.defensenews.com/air/2023/03/08/us-air-force-eyes-fleet-of-1000-drone-wingmen-as-planning-accelerates/

Harper, J. (2023b, March 20). Air Force plans to spend more than $6B on CCA drone programs over the next 5 years. *DefenseScoop*. https://defensescoop.com/2023/03/20/air-force-plans-to-spend-more-than-6b- on-cca-drone-programs-over-the-next-5-years/

Hitchens, T. (2020, October 15). Army aims to field TITAN terminals for all-domain ops in 2024. *Breaking Defense*. https://breakingdefense.com/2020/10/army-aims-to-field-titan-terminals-for-all-domain-ops-in-2024/

Interim report. (2023). Commission on Planning, Programming, Budgeting, and Execution Reform. https://ppbereform.senate.gov/wp-content/uploads/2023/08/PPBE-Commission-Interim-Report-Final.pdf

MacGregor, M., Modigliani, P., & Grant, G. (2022). *The pillars of the modern defense budgeting system for the Planning, Programming, Budgeting, and Execution (PPBE) Commission* (The Modern Defense Budgeting Series). MITRE Center for Data-Driven Policy. https://www.mitre.org/sites/default/files/2022-02/pr-20-03247-2-pillars-of-the-modern-defense-budgeting-system.pdf

McGarry, B. W. (2022). *DOD planning, programming, budgeting, and execution (PPBE): Overview and selected issues for Congress* (R47178). Congressional Research Service. https://crsreports.congress.gov/product/pdf/R/R47178

Middleton, M. W. (2006). *Assessing the value of the Joint Rapid Acquisition Cell.* [Naval Postgraduate School]. https://core.ac.uk/download/pdf/36696207.pdf

Office of the Secretary of Defense. (2019). *Establishment of the Space Development Agency*. https://aerospace.csis.org/wp-content/uploads/2019/04/Establishing_SDA_DoD_Memorandum.pdf

Office of the Under Secretary of Defense (Comptroller)/Chief Financial Officer. (2023). *Defense budget overview: United States Department of Defense: Fiscal year 2024 budget request*. https://www.asafm.army.mil/Portals/72/Documents/BudgetMaterial/2024/pbr/Army%20FY%202024%20Budget% 20Overview%20Briefing.pdf

Sapolsky, H. M., Gholz, E., & Talmadge, C. (2017). *US defense politics: The origins of security policy*. Routledge.

Space Development Agency. (n.d.). *Who we are*. https://www.sda.mil/home/who-we-are/

Uncrewed maritime systems: Navy should improve its approach to maximize early investments (GAO-22-104567). (2022). Government Accountability Office. https://www.gao.gov/assets/gao-22-104567.pdf

U.S. Department of the Army. (2023). *Department of Defense fiscal year (FY) 2024 budget estimates, Army justification book volume 2b of 2: Research, development, test & evaluation*. https://www.asafm.army.mil/Portals/72/Documents/BudgetMaterial/2024/Base%20Budget/rdte/RDTE-Vol%202-Budget%20Activity%204B.pdf

Zoldi, D. M. K. (2023, June 22). Navy hopes to make big waves with unmanned vehicles. *Inside Unmanned Systems*. https://insideunmannedsystems.com/navy-hopes-to-make-big-waves-with-unmanned-vehicles/

APPENDIX

Paper	***The Efficacy of Optimized Government–Industry–Academia Co-Education for Major Weapon Systems Cost/Price Analysis and Contract Negotiations*** Kelley Poree, Naval Postgraduate School

The Efficacy of Optimized Government–Industry–Academia Co-Education for Major Weapon Systems Cost/Price Analysis and Contract Negotiations

Kelley Poree—is the Contract Management Area Chair for the Department of Defense Management, Naval Postgraduate School (NPS). Before joining NPS in 2019, Kelley completed a recalibration tour, where after retiring as the Deputy Chief of Contracts for the Fighter Bomber Directorate, Chief of Contracts for the Development and Integration Division, and Warranted Contracting Officer (Unlimited) in 2015 at Wright–Patterson Air Force Base, OH, he entered civil service as a Contract Specialist from 2015 to 2017. After that, he participated in the limited and unlimited warrant contracting officer boards from 2017 to 2019. This recalibration tour highlighted gaps between acquisition theory, practice, and opportunities to build bridges. Kelley holds an MBA with a concentration in strategic purchasing from NPS. He is also a third-year doctoral student at Indiana Wesleyan University and received the 2021 Future of Pricing Award. [kelley.poree@nps.edu]

Abstract

The success of delivering and transitioning major weapon systems capabilities at the speed of relevance relies on the collaborative cost/price analysis and contract negotiations process buyers and sellers use to facilitate speed-to-contract award. However, Werber et al. (2019) found that insufficient knowledge of industry operations, risk management, and limited opportunities to attend joint formal education and training events influenced buyer understanding of requirements, cost/price analysis, and contract negotiations (p.120). In response, Werber et al. (2019) identified government–industry collaboration in the form of co-education as a potentially innovative strategy within the ecosystem to minimize these variations and related knowledge gaps (p.120). This study, which was a collaborative effort between graduate students in the Department of Defense Management at the Naval Postgraduate School and buyers and sellers from a major weapon system program office in the Midwest, explored perceptions on the efficacy of co-education for major weapon systems cost/price analysis and contract negotiations to minimize these variations using a common language software package, ProPricer Government Edition (GE). These findings indicate that this approach enables an open and honest transfer of data, information, and knowledge, facilitated by the practical use of ProPricer GE, to support collaboration and innovation, enhancing trust in buyer–seller relationships.

Introduction

Defense Acquisition System (DAS) buyers and sellers often attempt to conduct major weapon systems cost/price analysis and sole-sourced contract negotiations at the speed of relevance across 75 Major Defense Acquisition Programs (Office of the Under Secretary of Defense, Comptroller/Chief Financial Officer, 2023, p. 5). However, in a dynamic 21st-century national security environment, the Department of Defense's (DoD's) legacy acquisition system is still too slow to be competitive, only incrementally innovative, and optimized to a peacetime cadence inconsistent with the current speed of the great power competition (Congressional Research Service [CRS], 2023, p. iii; Kotila et al., 2023, p. 1; Wong et al., 2022, p. ix). One reason for this suboptimal peacetime cadence is the variation in buyer and seller predecessor education, training, and practice before buyers and sellers conduct cost/price analysis and contract negotiations. DAS buyers and sellers typically undergo specialized training at the individual, group, or organizational level specific to the agency or attend industry conferences and professional workshops. These education, training, and practice activities are rarely joint, creating the conditions for buyers and sellers to enter major weapon systems cost/price analysis and contract negotiations with varying and often conflicting degrees of competence (Werber et al., 2019. p. 124). Werber et al. (2019) identified government–industry co-education as an innovative strategy to address these variations and knowledge gaps (p.120). The researchers also highlighted two specific areas of government–industry co-education—industry rotations and industry resources (i.e., participants, presenters, standards, etc.) in internal training and development—and recommended government–industry interactions earlier in career development stages (Werber et al., 2019, p. 120).

Government–Industry–Academia Co-Education

Government–industry–academia co-education (G-I-A Co-Ed) builds upon government–industry co-education (Werber et al., 2019, p. 20) by leveraging Etzkowitz's (2003) Triple Helix Theory. This theory suggests that the key to enhancing innovation in a knowledge-based society lies in the university–industry–government interaction (p. 295). Within this triple helix, the industry serves as the production hub, the government provides contractual relationships, and the university (academia) is the wellspring of new technology and knowledge (Etzkowitz, 2003, p. 235). In *The New Economics for Industry, Government, and Education*, Deming (2018) described this interaction as an opportunity for management to redefine traditional boundaries and better serve the system's aim (p. 37). According to Deming (2018), everyone—stockholders, suppliers, employees, and customers—benefits from an optimized system (Location, 447). Deming's (2018) perspective on optimization involves understanding the interdependencies within the defense acquisition ecosystem, knowledge about variations, and the notion that each stage should benefit from more effort than the next stage or step (p. 93). In the context of the major weapon cost/price analysis and contract negotiations process (Table 1), optimization necessitates a consensus on buyer and seller roles within the ecosystem and knowledge about variations in education, training, and practice across these twelve steps, with an emphasis on understanding the cumulative effect on the cost/price analysis and sole-source contract negotiations process timelines (Deming, 2018, pp. 63–68).

Table 1. Major Weapon Systems Cost/Price Analysis and Contract Negotiations Process

Steps	Activity
1	Requirements Planning
2	Release Draft Letter Request for Proposal (RFP)
3	Approve Program/Project
4	Release RFP
5	Receive Proposal
6	Conduct Fact/Finding and Develop Technical Evaluation
7	Complete Pre-Price Negotiation Memorandum (Cost/Price Analysis)
8	Receive Business Clearance
9	Conduct Contract Negotiations
10	Complete Final Price Negotiation Memorandum
11	Receive Contract Clearance Approval
12	Award Contract

The optimization process, thus, must consider how ecosystem domains (i.e., education, training, practice, and execution) interact and influence each other toward contributing to DAS performance outcomes (Deming, 2018, p. 65).

Drucker's Five Essential Statements

Drucker et al. (2015) expanded Deming's (2018) perspective on optimization for DAS buyers and sellers through five essential statements (Location, 264). The first question is *What is our mission?* This question involves understanding the current mission, challenges, opportunities, and a growth mindset to consider if the mission requires revision (p. 6). For Drucker et al. (2015), this question includes a self-assessment to analyze challenges and opportunities to determine desired outcomes and results (p. 9). This approach begins with the end in mind and then suggests actionable steps to get there (Drucker et al., 2015, p. 9). Therefore, major weapon system buyers and sellers must consider an optimized cadence that supports a dynamic 21st-century national security environment. The second question is *Who is the customer?* Embedded in this question is the requirement to identify primary and helping customers and how these customers change over time (Drucker et al., 2015, p. 18). In the context of major weapon systems cost/price analysis and contract negotiations, multiple stakeholders across multiple domains comprise primary customers. For example, primary customers include the buyers and sellers in the education, training, and practice domains. However, buyers and sellers then transition from primary customers in these domains to supporting customers in the execution domain—when buyers and sellers conduct cost/price analysis and negotiations to deliver capabilities to the warfighter, the primary customer. Accordingly, "customers are never static" (Drucker et al., 2015, p. 21), and the ecosystem domains, among other things, must account for primary customers, supporting customers, and inter-domain transitions.

A related third essential question is *What does the customer value?* According to Drucker et al. (2015), external and internal customers base their needs on the realities of the situation and will behave rationally based on the circumstances (p. 33). In addition to accounting for primary supporting customer inter-domain transitions, the ecosystem must also account for the associated changes in customer needs and values in the process (i.e., changing values across education, training, practice, and execution domains). The fourth question is *What are our results?* This question includes a consensus on defining results, establishing metrics, and deciding what to keep or remove (Drucker et al., 2015, p. 45). This question assists each domain in learning, self-correcting, and understanding the cumulative effect on the overall process (Drucker et al., 2015, p. 53). Conversely, when one domain measure is independent of the others, conditions for casual unsystematic observations exist, limiting the understanding of how one domain affects the other and, by extension, overlooking the cumulative effect on the system. Drucker et al.'s (2015) fifth question centers on *What is the plan?* This question involves five elements: (1) deciding to abandon what does not work, (2) strengthening what does work, (3) creating conditions for innovation, (4) taking risks, and (5) analyzing and studying an essential performance area (p. 65).

Enhancing Buyer–Seller Trust Through Experiential and Interactive Learning

Consistent with Drucker et al.'s five essential questions, Handfield (2019) identified multi-stakeholder relationships, real-time analytics, and shared innovation risk as foundational concepts to support enhancing buyer–seller trust relationships (p. 195). With velocity emerging as the outcome performance metric for defense acquisition ecosystem buyers and sellers, analytics must support trust across multiple stakeholders, and the contractual guidelines should support shared innovation risk (Handfield, 2019, p. 198). This departure from the traditional buyer–seller relationship highlights the need for a new form of governance (Henfield, 2019, p. 198). The new form of governance, therefore, must consider a comprehensive plan. One approach to this comprehensive plan is to begin with experiential learning in a G-I-A Co-Ed environment. Kolb's (1984) experiential learning cycle is a widely accepted foundational model for adult learning (Morris, 2020, p. 1064). This model supports a cyclical process of learning experiences involving

1. concrete experiences, which include new experiences without bias;
2. reflective observations, which emphasize reflection on experiences from multiple perspectives;
3. abstract conceptualizations, which underscore creating concepts that assist in synthesizing into logically sound theories; and
4. active experimentation, using the theories to make decisions and solve problems (p. 30).

Consistent with the benefits of Kolb's (1984) experiential learning theory, Poree (2023) found that 83 of 111 U.S. military graduate students who completed Naval Postgraduate School cost/price analysis and contract negotiation courses between Winter 2021 and Summer 2022 agreed that active experimentation with the complementary software platform ProPricer GE enhanced critical thinking and problem-solving skills (p. 441). ProPricer GE integration into Naval Postgraduate School courses was based on the development of the 2018 DoD Sole Source Streamlining Toolbox, which highlighted that a "significant number of contractors use ProPricer software application for proposal development and analysis" (p. 4). Moreover, Cooper (2022) noted that approximately 70% of the major weapon system contractors use ProPricer, with a limited number of government agencies using the complementary proposal analysis software ProPricer GE (pp. 1–2). Given the favorable results with U.S. military officers and civilian populations, Poree (2023) recommended that future researchers extend co-education in the classroom to buyers and sellers from the execution domain or mission area.

Methods

This section outlines the research design, participant selection, and data collection procedures to explore the efficacy of optimized G-I-A Co-Ed for significant weapon systems cost/price analysis and contract negotiations.

Research Design

The research design included a qualitative approach with triangulation based on the need to understand various perspectives and opinions (Bryman, 2016, p. 386). According to Yin (2018), the major strength of a case study is the opportunity to use different data sources (p. 126).

Participants

The study included a purposive sample of two government buyers, two industry sellers supporting a major weapon systems program office in the Midwest, and 27 graduate students enrolled in MN3320, Cost/Price Analysis, and MN3321, Contract Negotiations, at the Naval Postgraduate School in Monterey, CA. Participants were selected based on their willingness to participate in the study and graduate students enrolled in the courses as part of the 815 Master of Science in Contract Management Curriculum. The class also included video recordings for buyers and sellers actively engaged in the mission area. Table 2 shows the total class population and a class percentage breakout. Thirteen U.S. Army graduate students comprised 42% of the participants; nine U.S. Navy graduate students, 29%; three U.S. Marine Corps graduate students, 10%; two U.S. Air Force graduate students, 6%; and two government buyers and two industry sellers, 6% and 6%, respectively.

Table 2. G-I-A Co-Ed Participants

Participants	Total Population	Class Percentage
U.S Army	13	42%
U.S. Navy	9	29%
U.S. Marines	3	10%
U.S. Air Force	2	6%
Government Buyer	2	6%
Industry Seller	2	6%
Total Class Population	**31**	**100%**

Data Collection Procedures

A 20-statement survey captured the data to understand the perceptions and opinions on optimized G-I-A Co-Ed, using a 7-point Likert scale. The scale ranged from Strongly Disagree, Somewhat Disagree, Disagree, Undecided, Agree, Somewhat Agree, to Agree Strongly. The survey opened on March 4, 2025, and closed on March 29, 2024.

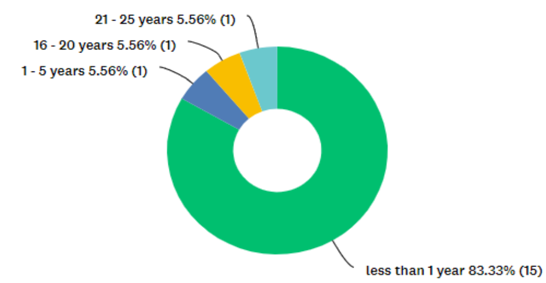
Results

The results centered around the primary research question: How do participants perceive the efficacy of optimized G-I-A Co-Ed to enhance buyer and seller high-trust collaboration and innovation? Table 3 shows a population of 31 academic student buyers and sellers, including those from the mission area. Eighteen responses produced a response rate of 58%, with 62% of academic student buyers, 54% of academic student sellers, 100% of government buyers, and 50% of industry sellers completing the survey.

Table 3. G-I-A Co-Ed Survey Response Rates

Population	Total Population	Responses	Response Rate
Academia Student/Buyers	13	8	62%
Academia Student/Sellers	14	7	50%
Government Buyers	2	2	100%
Industry Sellers	2	1	50%
Total	**31**	**18**	**58%**

As depicted in Figure 1, 15 (or 83%) of the participants had less than 1 year of major weapon system cost/price analysis and contract negotiation experience, one (or 5.56 %) had 1 to 5 years of experience in this domain, one (or 5.56%) had 16 to 20 years of experience, and one (or 5.56 %) had between 21 and 25 years of cost/price analysis and contract negotiations experience.



ANSWER CHOICES	RESPONSES	
less than 1 year	83.33%	15
1 - 5 years	5.56%	1
16 - 20 years	5.56%	1
21 - 25 years	5.56%	1
6 - 10 years	0.00%	0
11 - 15 years	0.00%	0
more than 25 years	0.00%	0
TOTAL		18

Figure 1. G-I-A Co-Ed Participant Years of Experience

Key Findings

Figure 2 shows the survey results for the 17 Likert scale statements, emphasizing Agreed, Somewhat Agreed, and Strongly Agreed as the primary outcomes. Statements 1 and 2 centered on participant type and years of experience and, therefore, were omitted. These results were categorized into three key findings that support the efficacy of optimized G-I-A Co-Ed for major weapon systems cost/price analysis and contract negotiations.

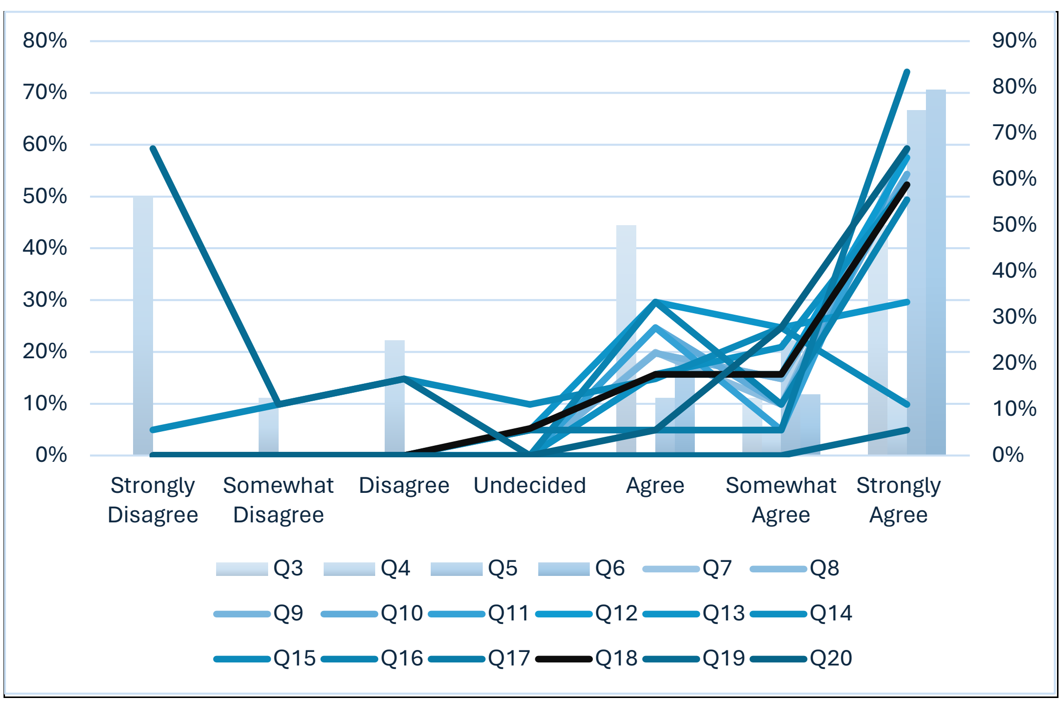


Figure 2. G-I-A Co-Ed Survey Results

Variations Exist in Buyer–Seller Education, Training, and Practice Domains

- **Statement 3.** 44% of the respondents Strongly Agreed, 11 % Somewhat Agreed, and 44% Agreed that variations exist in buyer and seller education, training, and practice for major weapon system cost/price analysis and contract negotiations.
- **Statement 4.** 50% Strongly Disagreed, 11% Somewhat Agreed, and 22% Disagreed that education, training, and practice variations DO NOT negatively affect buyer and seller abilities to conduct major weapon system cost/price analysis and contract negotiations.

G-I-A Co-Ed Minimizes Variations in Buyer–Seller Education, Training, and Practice

- **Statement 5.** 67% of the respondents Strongly Agreed, 22% Somewhat Agreed, and 11% Agreed that G-I-A Co-Ed provides insight into buyer and seller motivations, operations, and perspectives on cost, schedule, and performance risks.
- **Statement 6**. 71% of the respondents Strongly Agreed, 12% Somewhat Agreed, and 18% Agreed that active experimentation with ProPricer GE in G-I-A Co-Ed enables work traceability and a systematic approach to analyzing work breakdown structures, tasks, and the associated basis of estimates.
- **Statement 7.** 61% of the respondents Strongly Agreed, 17% Somewhat Agreed, and 22% Agreed that active experimentation with ProPricer GE in G-I-A Co-Ed enables a systematic approach to fact-finding.
- **Statement 8.** 67% of the respondents Strongly Agreed, 11% Somewhat Agreed, and 22% Agreed that using ProPricer GE in the G-I-A Co-Ed enables a systematic approach to establishing minimum, objective, and maximum positions.
- **Statement 9.** 61% of the respondents Strongly Agreed, 11% Somewhat Agreed, and 22% Agreed that active experimentation with ProPricer GE in G-I-A Co-Ed enables a systematic approach to provide offers and counteroffers in the negotiations process.
- **Statement 10.** 61% of the respondents Strongly Agreed, 11% Somewhat Agreed, and 28% Agreed that active experimentation with ProPricer GE in G-I-A Co-Ed enhanced understanding of fair and reasonable determinations for buyers and sellers.
- **Statement 11.** 67% of the respondents Strongly Agreed, 6% Somewhat Agreed, and 28% Agreed that participating in G-I-A Co-Ed earlier in the buyer and seller professional development process could increase individual competence in major weapon systems cost/price analysis and contract negotiations.

G-I-A Co-Ed Enhances Buyer and Seller Trust, Collaboration, and Innovation

- **Statement 12.** 65% of the respondents Strongly Agreed, 18% Somewhat Agreed, and 18% Agreed that participating in G-I-A Co-Ed earlier in the buyer and seller professional development process could increase the organizational capability to deliver major weapon systems on time and within budget.
- **Statement 13.** 33% of the respondents Strongly Agreed, 28% Somewhat Agreed, 33% Agreed, and 5.56% were Undecided on whether participating in G-I-A Co-Ed enhances trust in the buyer–seller relationship required to deliver warfighter capabilities.
- **Statement 14.** 59% of the respondents Strongly Agreed, 24% Somewhat Agreed, and 18% Agreed that participating in G-I-A Co-Ed (and using ProPricer GE) supports an open and honest transfer of data, information, and knowledge.
- **Statement 15**. 11% of the respondents Strongly Agreed, 28% Somewhat Agreed, 17% Agreed, 11% were undecided, 17% Disagreed, 11% Somewhat Disagreed, and 6% Strongly Disagreed that using ProPricer GE in a G-I-A Co-Ed context limits the ability of buyers and sellers to act opportunistically.

- **Statement 16.** 56% of the participants Strongly Agreed, 11% Somewhat Agreed, and 33% Agreed that participating in G-I-A Co-Ed improves understanding of buyer and seller motivations, behaviors, and trends that shape intelligent business decisions.
- **Statement 17.** 83% of the participants Strongly Agreed, 6% Somewhat Agreed, 6% Agreed, and 6% were Undecided on the extent to which exposure to buyers and sellers from the mission area enhanced understanding of the challenges associated with leading major weapon systems buyers and selling organizations in a dynamic 21st-century national security environment.
- **Statement 18.** 59% Strongly Agreed, 18% Somewhat Agreed, 18% Agreed, and 6% were Undecided on whether participation in G-I-A Co-Ed creates the conditions for forging and enhancing trust relationships between buyers and sellers.
- **Statement 19.** 67% Strongly Disagreed, 11% Somewhat Disagreed, 17% Disagreed, and 6% Agreed that trust IS NOT essential for buyers and sellers conducting major weapon system cost/price analysis in a dynamic 21st-century national security environment.
- **Statement 20.** 67% of respondents Strongly Disagreed, 28% Somewhat Disagreed, and 6% Agreed that G-I-A Co-Ed creates the environment for buyer/seller collaboration and conditions for innovation.

Secondary measures included student course evaluation form (CEF) scores and comments in Table 4 and Table 5. Table 4 shows a total response rate of 100% (27 of 27 through frequencies of 24, 2, and 1) related to student perspectives on course learning and a response rate of 100% for 23 and 4 related to student perspectives on course content and design.

Table 4. Academia Student Buyer and Student Seller Course Evaluation Form Scores

Question: What you learned in the course	MAX	MIN	AVG	STDEV	VAR	Frequency of 5 Rating	Frequency of 4 Rating	Frequency of 3 Rating	Frequency of 2 Rating	Frequency of 1 Rating
1.1. I developed new skills and abilities.	5	3	4.81	0.48	0.23	23	3	1	0	0
1.2. I improved my understanding of the subject.	5	4	4.89	0.32	0.1	24	3	0	0	0
1.3. I strengthened my analytic capabilities.	5	4	4.93	0.27	0.07	25	2	0	0	0
1.4. I enhanced my ability to think critically.	5	3	4.85	0.46	0.21	24	2	1	0	0
1.5. Overall, I learned a great deal.	**5**	**3**	**4.9**	**0.46**	**0.2**	**24**	**2**	**1**	**0**	**0**
Question: Content and design of the course	MAX	MIN	AVG	STDEV	VAR	Frequency of 5 Rating	Frequency of 4 Rating	Frequency of 3 Rating	Frequency of 2 Rating	Frequency of 1 Rating
2.1. The course material engaged me in the subject matter.	5	4	4.81	0.4	0.16	22	5	0	0	0
2.2. The course assignments reinforced course content.	5	5	5	0	0	27	0	0	0	0
2.3. The course content was relevant to my program of study.	5	4	4.96	0.19	0.04	26	1	0	0	0
2.4. This course was academically challenging.	5	3	4.59	0.57	0.33	17	9	1	0	0
2.5. Overall, the course was well designed.	**5**	**4**	**4.9**	**0.36**	**0.1**	**23**	**4**	**0**	**0**	**0**

Overall, the results revealed that student buyers and student sellers developed new skills and abilities and improved their understanding of the concepts and activities associated with major weapon systems cost/price analysis and contract negotiation, with scores of 4.82 / 5.00 (or 96%) and 4.89 / 5.00 (or 98%), respectively. Regarding course design, respondents scored the relevance of the course content to the program of study 4.96 / 5.00 (or 99%).

Table 5 captures 10 student-related comments.

Table 5. Participant Comments

Number	Comments
1	"The course was challenging and rewarding. It provided real-world experiences and points of view from civilians currently in the work field. I learned a lot that I can apply in my career field."
2	"Incorporation of industry and external acquisition professionals provides unique insight into the challenges we face outside the classroom. Using new and innovative contract pricing tools (ProPricer GE) was enlightening, and I saw that progress can be made in efficiency and effectiveness."
3	"Excellent interactivity with the class; the course is well designed to promote learning by doing."
4	"The course was well designed to integrate government and industry in the educational setting to better prepare students for the realities of the mission."
5	"The co-education between the government and seller representatives was beneficial."
6	"Integrating software to the academic environment."
7	"Industry partner presence. Choice of case studies. Tutorial Support. Take home lab assignments. Group assignments"
8	"The course was a good blend of student experience, industry inputs and point of views in the form of guest appearance, and customized course content lead by Prof. Poree."
9	"Continue implementing guest attendance for real-world civilians. Their perspective was beneficial for altering the government's mindset and point of view. Pro Pricer was a great program to practice with and learn."
10	"Interacting with ProPricer and industry partners was eye-opening. Negotiating among classmates was a great learning experience. Seeing how two groups reached different outcomes (yet still sealed the deal) highlights the complexity of issues we will face when we return to the field."
11	"I watched the last lecture, during which you demonstrated ProPricer with the IGCE and seller's proposal. ProPricer would have been awesome when I was a buyer a/o PCO! It makes everything SO MUCH EASIER!"
12	"Buyers and sellers have different education, training, and practice paths."

Discussion

The efficacy of optimized G-I-A Co-Ed for major weapon systems cost/price analysis and contract negotiations survey result, associated CEF scores, and related comments center on overarching themes: (1) variations exist in education, training, and practice domains; (2), G-I-A Co-Ed minimizes variations in education, training, and practice domains; and (3) G-I-A Co-Ed enhances buyer–seller trust, collaboration, and innovation. While a promising first step, readers should cautiously interpret findings based on (1) the limited number of government and industry participants and (2) the scope of this study, which includes Steps 5 to 12 and not Steps 1 to 4 in Table 1.

The top section of Figure 3 depicts variations in education, training, practice, and execution in the "as-is" in major weapon systems buyer and seller cost/price analysis and contract negotiation. The bottom section captures the optimized G-I-A Co-Ed MN3320/21 Cost/Price Analysis and Contract Negotiations.

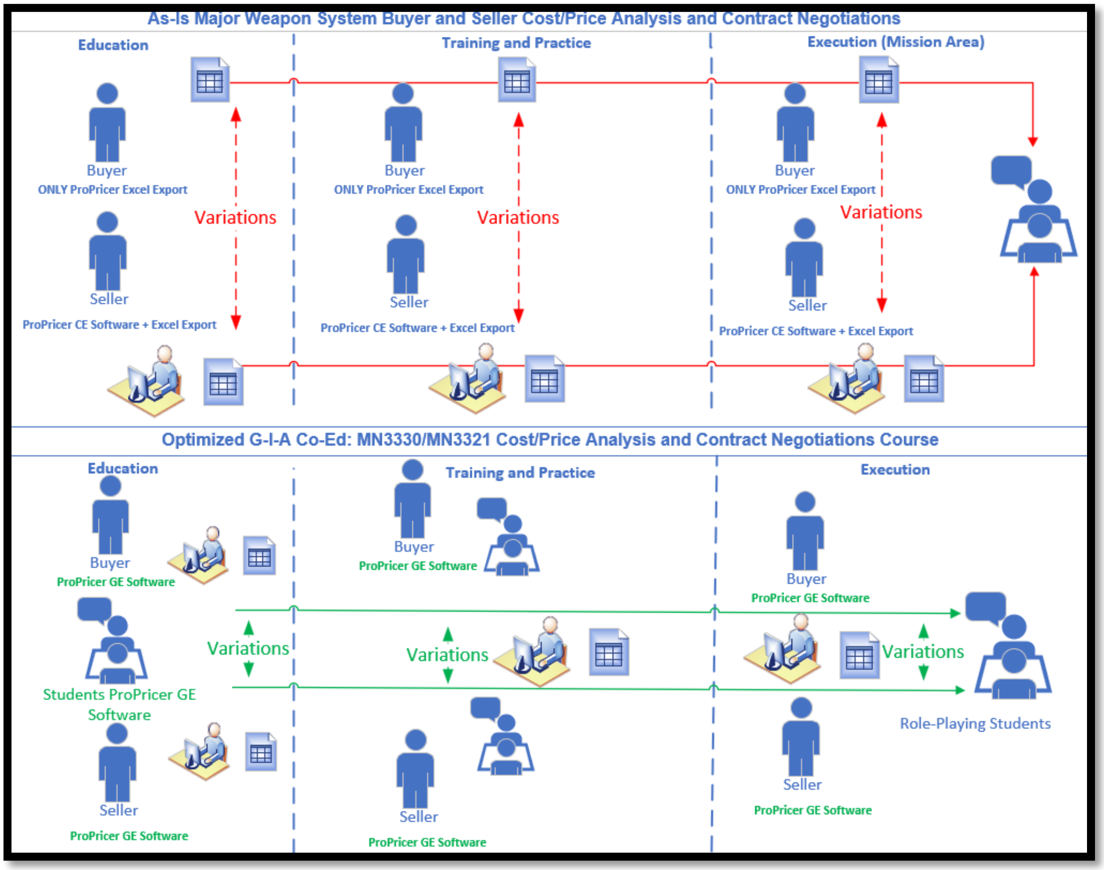


Figure 3. Optimized G-I-A Co-Ed for Cost/Price Analysis and Contract Negotiations

Variations Exist in Buyer-Seller Education, Training, and Practice

As depicted in Figure 4, the results from Statements 3 and 4 support that variations exist in buyer and seller education, training, and practice. Expressly, in Statement 3, 44% of the respondents Strongly Agreed, 11% Somewhat Agreed, and 44% Agreed that variations exist in buyer and seller education, training, and practice. For Statement 4, 50% of participants Strongly Disagreed and 11% Somewhat Disagreed that education, training, and practice variations do not negatively affect buyer and seller abilities to conduct major weapon systems cost/price and contract negotiations. These results underscore the relationship between variations in the education, training, and practice domains and the impact these variations have in the execution domain. These results are consistent with Werber et al.'s (2019) findings regarding buyers possessing insufficient knowledge of industry operations, risk management, and limited opportunities to attend joint formal education and training events influenced their understanding of requirements, cost/price analysis, and contract negotiations (p. 120). Moreover, the results are consistent with Deming's (2019) perspective on understanding interdependencies and variations (p. 93). For Drucker et al. (2015), these results also provide self-assessment results from which to consider the plan toward optimization (p. 9).

G-I-A Co-Ed: Consistent and Systematic Buyer–Seller Education, Training, and Practice

As captured in Figure 1, Statements 5 to 11 results support G-I-A Co-Ed and provide a more consistent and systematic approach in education, training, and practice for major weapon systems cost/price analysis and contract negotiations. For example, 67% of the respondents Strongly Agreed, 22% Somewhat Agreed, and 11% Agreed that G-I-A Co-Ed provides insight into buyer and seller motivations, operations, and perspectives on cost and schedule performance risks (Statement 5). When combined with the results of Statement 6, where 70% of the respondents Strongly Agreed, 11.76% Somewhat Agreed, and 17.65% Agreed that active experimentation with ProPricer GE enabled work traceability and a systematic approach to analyzing work breakdown structures, tasks, and associated basis of estimates, this supports Handfield's (2019) position on the importance of real-time analytics to support multi-stakeholder relationships (p. 195). Strong agreements across Statements 8 to 11 also support a collective understanding of major weapon system cost/price analysis and contract negotiations process for major weapon systems (i.e., beginning with Step 5 through Step 12 of Table 1).

Additional qualitative student statements support that G-I-A Co-Ed minimizes buyer–seller education, training, and practice variations. For example, participant responses such as these support a common understanding across different populations with different competency levels:

- "Incorporation of industry and external acquisition professionals provides unique insight into the challenges we face outside the classroom. The use of new and innovative contract pricing tools (ProPricer GE) was enlightening in seeing that progress can be made in efficiency and effectiveness" (participant response, number 2).
- "Interacting with ProPricer as well as industry partners was eye-opening. The process of negotiating among classmates was a great learning experience, and seeing how two groups reached different outcomes (yet still sealed the deal) highlights the complexity of issues we will face when we go back to the field" (participant response, number 10).

G-I-A Co-Ed Enhances Buyer and Seller Trust, Collaboration, and Innovation

Overall, strong agreement across Statements 12, 14, 16, 17, 18, 19, and 20 support the notion that G-I-A Co-Ed enhances buyer–seller trust, collaboration, and innovation. Respondents strongly agreed on the benefits of early participation in G-I-A Co-Ed, open and transparent data transfer, understanding buyer and seller motivations, and creating the conditions to enhance trust, collaboration, and innovation. These results are consistent with Drucker et al.'s (2015) view on the need for leaders to create the conditions for innovation, take risks, and analyze and study essential performance areas (p. 95). Handfield (2019) extended Drucker et al.'s (2015) viewpoint by underscoring the importance of sharing innovation risk and real-time analytics that enhance buyer–seller trust.

Conclusion and Recommendations

While variations in buyer and seller education, training, and practice domains exist, results from this study provided insight into the efficacy of optimized G-I-A Co-Ed for major weapon systems cost/price analysis and contract negotiation. Specifically, when buyers and sellers use near real-time analytics with ProPricer GE in the sole-source contracting process, participants with varying degrees of experience and competence benefit from concrete experiences, reflective observations, abstract conceptualizations, and active experimentation earlier in the buyer and seller professional development process. The success of delivering and transitioning major weapon systems capabilities at the speed of relevance, thus, relies on the integrated and synchronized G-I-A interactions. These interactions, in part, facilitate the speed-to-contract award and, by extension, a major weapon systems cost/price analysis and contract negotiations cadence consistent with the needs of a dynamic 21st-century national security environment.

The study also generated three recommendations for future researchers to consider within the defense acquisition ecosystem. First, researchers should expand future cost/price analysis and contract negotiation studies to include Steps 1 to 4 of the process in Table 1: requirements planning, release draft RFP, approval program, and release RFP, respectively. Establishing a baseline of a buying organization's existing baseline for cost/price analysis and contract negotiations without ProPricer GE and then measuring the integration of ProPricer GE against the baseline might provide additional insights into G-I-A Co-Ed impacts on the corresponding personnel costs. Second, researchers should study adding more buyers and sellers from the mission area into future courses to provide a more comprehensive outcome. Third, future researchers could extend Deming's (2018) perspective on how stockholders, suppliers, employees, and customers benefit from an optimized system that includes subcontractors and suppliers who use ProPricer Contractor Education.

References

Bryman, A. (2016). Social research methods. Oxford University Press.

Congressional Research Service. (2023). Defense acquisitions: How DOD acquires weapon systems and recent efforts to reform the process. https://crsreports.congress.gov

Cooper, Z. (2022). Perceptions on the feasibility of implementing innovative cost and price analysis software across Naval Sea Systems Command. Acquisition Research Program.

Deming, W. E. (2018). The new economics for industry, government, and education. MIT Press.

Drucker, P. F., Hesselbein, F., & Kuhl, J. S. (2015). Peter Drucker's five most important questions: Enduring wisdom for today's leaders. John Wiley & Sons.

Etzkowitz, H. (2003). Innovation in innovation: The triple helix of university–industry–government relations. Social Science Information, 42(3), 293–337.

Handfield, R. (2019). Shifts in buyer-seller relationships: A retrospective on. Industrial Marketing Management, 83(5), 194–206.

Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development. Prentice-Hall

Kotila, B., Drezner, J. A., Bartels, E. M., Hill, D., Hodgson, Q. E., Huilgol, S. S., Manuel, S., Simpson, M., & Wong, J. P. (2023). Strengthening the defense innovation ecosystem.

Morris, T. H. (2020). Experiential learning: A systematic review and revision of Kolb's model. Interactive Learning Environments, 28(8), 1064–1077. https://doi.org/10.1080/10494820.2019.1570279

Office of the Under Secretary of Defense, Comptroller/Chief Financial Officer. (2023, March 5). Program acquisition cost by weapon system. https://comptroller.defense.gov/Budget-Materials/Budget2024/

Poree, K. (2023). Educational leadership, collaboration, and relevance: A get real, get better approach to innovating major weapon systems cost/price analysis and contract negotiations courses in higher education. Acquisition Research Program.

Werber, L., Ausink, J., Daugherty, L., & Phillips, B. (2019). An assessment of gaps in business understanding and knowledge of industry within the defense acquisition workforce. RAND. https://www.rand.org/pubs/research_reports/RR2825.html

Wong, J. P., Younossi, O., LaCoste, C. K., Antón, P. S., Vick, A., Weichenberg, G., & Whitmore, T. C. (2022). Improving defense acquisition insights from three decades of RAND research. RAND.

Yin, R. K. (2018). Case study research and applications (Vol. 6). Sage

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net