SYM-AM-25-300

Proceedings
of the
Twenty-Second Annual Acquisition Research Symposium and Innovation Summit

Volume III
Accepted Papers Not Presented

Creating Synergy for Informed Change: Transitioning Technology to the Warfighter

May 7–8, 2025

Published: May 5, 2025

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

Preface & Acknowledgements

The purpose of the “Creating Synergy for Informed Change: Transitioning Technology to the Warfighter, 22nd NPS Annual Acquisition Research Symposium and Innovation Summit” is to provide a forum for the presentation of scholarly acquisition research, as well as for dialogue between scholars and acquisition policy-makers and practitioners. Research papers and presentations are given on recently completed and on-going Departments of Defense and US Navy (DoD/DON)-sponsored projects conducted by researchers at a variety of research institutions. Senior DoD/DON acquisition officials serve as panelists or keynote speakers to present their critiques and comments on research papers and priorities.

This year our symposium is coupled with an Innovation Summit and takes up the theme of “Transitioning Technology.” The goal of this dual event is to explore and promote innovative ways to transition technology from research and development to programs of record to support the warfighter.

Although attendees come from many U.S. locations, as well as from some international locales, a large number are from Naval Postgraduate School (NPS) where faculty members and graduate students engage in acquisition-related research. In particular, NPS graduate students are an integral component of research and dialogue. The Symposium serves an essential part of their graduate learning experience and provides them with the opportunity to meet with senior policymakers, practitioners, and distinguished scholars.

We gratefully acknowledge the ongoing support and leadership of our sponsors, whose foresight and vision have assured the continuing success of the Acquisition Research Program:

- Assistant Secretary of the Navy for Research, Development and Acquisition (ASN(RDA))
- Program Executive Officer, Integrated Warfare Systems
- Director, Acquisition Career Management, U.S. Navy
- Office of the Under Secretary of Defense (Acquisition & Sustainment)

The research presented in this report was supported by the Acquisition Research Program, Graduate School 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; and Acquisition 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

Timing is Everything: Schedules, Models, and Analysis

Charles Pickar—retired from the Naval Postgraduate School (NPS) faculty, where he taught project management, defense acquisition, and systems engineering. Previously, he led the Applied Systems Engineering Program Area at Johns Hopkins University Applied Physics Laboratory. He is a retired Army officer with extensive experience in the U.S. defense industry, including Director and VP levels at Lockheed Martin, Northrop Grumman, and SAIC. He has also served as Chair of the Systems Education Technical Committee of the IEEE Systems Council. His research and published work focus on applying systems engineering and system dynamics analytical approaches to defense acquisition problems. [ckpickar@gmail.com]

Raymond Franck—retired from the NPS faculty in 2012. He retired from the Air Force in 2000 in the grade of Brigadier General. His active-duty career included operational tours, staff positions, leadership assignments, and, finally, head of the Department of Economics and Geography, United States Air Force Academy. His published work includes journal articles, book chapters, and research reports on military innovation and defense acquisition management. [cfranck215@aol.com]

Abstract

The complexity and scale of defense projects necessitate innovative project management and scheduling approaches. Digital twins, a digital representation of physical entities, transform how projects are planned, executed, and monitored. This paper explores the definition, applications, advantages, and challenges of implementing digital twins in project management. Additionally, integrating artificial intelligence (AI) predictive delay analysis processes provides an advanced framework for optimizing execution and risk mitigation. This paper examines (a) how real-time digital replicas and AI-driven predictive analytics using defense acquisition data can enhance decision-making, efficiency, and project outcomes in defense project management and (b) how prediction markets might enhance the timeliness and quality of information for program management—leading to better program outcomes. While the technical advances are impressive, they rely on information. Also, program management involves human skills and knowledge. Prediction markets have demonstrated promising capabilities to provide timely and accurate information for program management—with or without state-of-the-art technical means.

Digital Twins and Supporting Tools

Introduction

Project management requires tools and methods to plan, execute, and monitor progress. The advent of digital twins (DTs), technical advancements like artificial intelligence (AI), and techniques like predictive analytics offer modern alternatives to traditional tools.

Schedules outline the planned sequence and duration of weapon system development events. When accurate, they offer warfighters a dependable delivery date. However, programs often miss these deadlines, breaking warfighter trust and leading to cost and performance issues. Defense project management involves intricate planning, execution, and monitoring to meet stringent requirements. Traditional methodologies struggle to address unforeseen challenges, resource constraints, and delays effectively. A DT provides a virtual (and simulated) replica of a physical project (weapon system development), allowing project managers to simulate, analyze, predict, and execute outcomes in real time. The convergence of DTs, AI, and predictive delay analysis (PDA) offers an innovative paradigm shift, leveraging real-time acquisition data and predictive insights to enhance project execution.

This paper explores the feasibility of utilizing AI and DT approaches to offer a model for creating and executing project schedules. By leveraging these advanced technologies, the study aims to enhance accuracy, efficiency, and adaptability in project scheduling processes.

The rise of DTs and AI marks a new era for project management, particularly in schedule management and the potential for improving the schedule estimating and executing process. Instead of using virtual replicas of the system being developed, we suggest using a DT as a model or simulation of the developed weapon system. With AI's analytical power, this idea can improve project timelines, predict delays, and boost efficiency. DTs serve as the activity monitoring tool. PDA using the schedule delay factor data supported by acquisition data is incorporated into the AI. AI then monitors the DT's inputs and alerts the decision-maker to the need for action. Predicting schedule problems is the end state.

This effort is part of a continuing research agenda started by Franck et al. (2016). The latest paper in this line of research (Franck & Pickar*, 2024*) discussed complexities in the estimation and execution dynamic processes that determine success or failure in schedule execution and the "wisdom of crowds" and prediction markets concepts and relationship to schedule and program management decisions (Frank & Pickar, 2024).

Literature Review

This study explores using DTs as an environment for practicing project management decision-making. The tools we propose to assist in this effort are DTs as a simulation environment and PDA and machine learning (ML) as a subset of AI. The literature sources include *inter alia* defense-focused papers and systems research on managing defense projects. An unexpected resource for studying how DTs and delay analytics can be used in defense project management has been the construction industry, where practitioners have explored the use of DTs, PDA, and AI for some time.

Grieves initially proposed the concept of DTs in the context of Product Life Cycle Management (PLM) in 2002 (Grieves, 2002). DTs were originally digital replicas of physical assets. They have since evolved into models of non-physical systems and processes. Grieves and Vickers believe DTs offer a dynamic approach to managing the development of complex systems (Grieves & Vickers, 2017). In project management, DT models can replicate project components, facilitating improved planning, execution, and monitoring. To manage weapon system development execution, DTs can provide a real-time linkage between the work accomplished and the project management schedule (and plan), providing the PM insight into favorable and unfavorable execution developments. Figure 1. shows the basic model (modified to add the project management function; Grieves, 2002.

Figure 1. Project Management DT Environment
(Grieves, 2002)

DTs were initially used in the aerospace sector, where NASA used mirrored simulation models for system maintenance and failure prediction (Uhlenkamp et al., 2019). Over time, DTs have evolved to integrate the Internet of Things (IoT) sensors, AI, ML, and data analytics, providing real-time insights into the development of the system and enabling life cycle management (Attaran & Celik, 2023; Pan & Limao, 2021).

Project schedules are critical for completing projects, as they outline the timeline of tasks and activities. DT models can enhance project scheduling by providing real-time data and predictive analytics. Project managers can identify potential delays by simulating different scenarios and adjusting schedules accordingly (Tao et al., 2019). DTs can provide a data-driven decision-making tool by continuously analyzing project performance. This real-time monitoring provides better forecasting of project timelines and improved resource allocation by understanding workload distribution, enabling proactive real-time adjustments. A DT also simulates real-world scenarios for the project process environment, providing a safe environment for exploring solutions to project problems.

Predictive Delay Analysis and Machine Learning

ML and PDA are distinct yet complementary concepts within data-driven decision-making. ML is a subfield of AI that involves the development of algorithms capable of identifying patterns in data and making predictions or decisions without explicit programming (Mitchell, 1997). These algorithms, including regression models, decision trees, support vector machines, and neural networks, have been widely used in healthcare, finance, and manufacturing due to their flexibility and predictive power.

In contrast, PDA is a specialized technique within project management, predominantly in the construction and engineering sectors. Its primary objective is to forecast project delays by analyzing scheduling data—typically derived from critical path method (CPM) schedules—alongside historical and real-time performance metrics (Arditi & Pattanakitchamroon, 2006). PDA relies on various methods such as trend analysis, rule-based systems, Monte Carlo simulations, and, increasingly, ML techniques.

Although PDA is domain-specific, ML serves as the general-purpose analytical thread. The intersection of the two arises when ML is used to improve the accuracy and automation of delay predictions. For example, supervised ML models can be trained on historical project data to identify patterns in activity durations, programmatic issues, procurement timelines, or subcontractor performance indicative of future delays.

The difference between PDA and ML lies in their scope and application**.** While ML offers a broad methodological framework applicable to numerous domains, PDA is a targeted application for temporal risk and delay forecasting in projects. Integrating ML into predictive delay analysis represents a convergence of generalizable algorithmic intelligence with domain-specific scheduling insights, potentially yielding more data-informed project risk assessments (Marzouk & El-Rasas, 2014).

Classical scheduling methods require manual estimation and static modeling of task durations and dependencies. These approaches lack the flexibility to respond to real-time changes and uncertainties, leading to inaccurate timelines and inefficient resource use (Kerzner, 2022). Leveraging ML and AI, recent studies demonstrate a growing shift from traditional reactive approaches to predictive, data-driven methodologies. PDA traces its origins to traditional project scheduling and risk management methodologies, including Critical Path Method (CPM) and Earned Value Management (EVM). A PDA model typically uses ML, statistical modeling, or simulation to identify the probability and impact of delays before they occur.

Table 1. Core Components of PDA

Component	Description
Input Data	Project schedules, progress reports, weather data, resource logs, risk registers, subcontractor performance, etc.
Features	Task duration, lag times, critical path activities, manpower fluctuations, change orders, etc.
Modeling Technique	Regression, Decision Trees, Random Forests, Bayesian models, or Deep Learning for complex dependencies
Output	Probability of delay per task, delay forecasts, and risk classification (low/medium/high)
Visualization	Gantt overlays, risk heatmaps, delay likelihood timelines

Table 1 lists the components of PDA. Of note are the input and output components, which will be included in the model suggested later in this paper.

A PDA model comprises the systematic collection and integration of data, which consists of historical schedule performance metrics, explicitly contrasting Planned versus Actual outcomes, and documentation of change orders and incidents of rework. The appropriate analytical technique is determined upon the compilation and preparation of this data. Potential methodologies include logistic regression for estimating the likelihood of delays or applying random forest algorithms to enhance interpretability and address nonlinear relationships (Ghimire & Mishra, 2019). The deployment of the delay analysis model is intended for integration within contractor-available project management software platforms, such as Primavera, MS Project, or JIRA, with the capacity for real-time updates through the incorporation of progress feeds throughout the execution phase. Different established visualization tools, including Power BI and Tableau, could be utilized to present real-time forecasts and associated risks to the critical path. Potential use cases include detecting likely delays in hardware delivery due to supplier risk and forecasting design freeze violations on critical components. Using feedback loops from the DT to auto-update the schedule would keep the PM in the loop (Liu et al., 2021).

PDA supports proactive project management by providing an opportunity to anticipate potential delays (Lee, 2017). It can also improve the accuracy of project timelines, making it easier to set realistic deadlines. The awareness of when and where delays might occur allows managers to allocate resources more effectively to avoid chokepoints.

Gondia et al. examined the use of supervised ML algorithms—including Support Vector Machines, Decision Trees, and Random Forests—to predict construction project delays by modeling the intricate interdependencies of delay risk sources (Gondia et al., 2020). The paper finds that multifactorial and interrelated risks often cause construction delays that traditional methods struggle to predict accurately. The study further reveals that ML models can effectively capture nonlinear relationships between variables and deliver more reliable and proactive delay risk assessments. Notably, the Random Forest algorithm demonstrated superior predictive accuracy and interpretability performance.

Machine Learning Algorithms in Classification and Regression

ML models have proven indispensable tools in modern PDA. These models, particularly those employing supervised learning algorithms, can analyze vast amounts of data to identify patterns and accurately predict outcomes. Support Vector Machines (SVMs), Decision Trees, and Random Forests are the most notable algorithms used for this purpose. Each algorithm presents unique characteristics regarding modeling capacity, interpretability, and performance.

Support Vector Machines

SVMs are supervised learning models primarily employed for classification tasks, although they can also be adapted for regression problems. The core principle of SVMs is to determine an optimal hyperplane that separates data points of different classes with the maximum possible margin. The data points closest to the decision boundary, termed *support vectors*, are critical in defining this margin (Cortes & Vapnik, 1995).

SVMs are particularly effective in high-dimensional spaces and are well-suited to problems where the number of features exceeds the number of observations. In cases where the data is not linearly separable, the SVM utilizes a *kernel trick* to project the data into a higher-dimensional space where a linear separation becomes feasible (Schölkopf et al., 1998). Common kernels include linear, polynomial, radial basis function (RBF), and sigmoid.

Decision Trees

Decision Trees are versatile, non-parametric models capable of performing both classification and regression tasks. They operate by recursively partitioning the input space into subsets based on feature values, forming a tree-like structure of decision rules (Quinlan, 1986). At each internal node, the algorithm selects a feature and a corresponding threshold that best splits the data to increase *purity*—a measure of how homogenous the resulting subsets are to the target variable.

The interpretability of decision trees is a significant advantage, as the resulting model can be easily visualized and understood. However, decision trees are prone to overfitting, especially when they grow deep and complex. Techniques such as pruning, setting maximum depth, or limiting the number of samples per leaf are commonly used to mitigate overfitting (Loh, 2011).

Random Forests

Random Forests is a learning method that enhances the performance of decision trees by constructing a multitude of trees and aggregating their outputs (forest). This approach, known as *bagging* (bootstrap aggregating), involves training each tree on a random subset of the data and selecting random subsets of features at each split, thereby introducing diversity among the trees (Breiman, 2001).

For classification tasks, the final output is determined by a majority vote among the individual trees, whereas for regression, the output is the average of the predictions. Random Forests offer improved generalization performance over individual decision trees, primarily due to their ability to reduce variance and avoid overfitting. Although Random Forests are less interpretable than single decision trees, they are generally more accurate and robust across various tasks. The trade-off lies in increased computational requirements and reduced transparency. These algorithms form a foundational component of many ML pipelines and continue to be extensively applied across various fields, including bioinformatics, finance, and engineering applications.

Table 2. ML Algorithm Summary Comparison

Algorithm	Application	Strengths	Limitations
SVM	Classification / Regression	Effective in high-dimensional spaces; robust with clear margin separation	Computationally intensive; sensitive to parameter tuning
Decision Trees	Classification / Regression	Simple to understand and interpret; fast training	Prone to overfitting; high variance
Random Forests	Classification / Regression	High accuracy; reduces overfitting; robust to noise	Less interpretable; increased training time

Key challenges for all ML algorithms include poor data quality, lack of project standardization, and resistance to adopting opaque "black-box" models. Many ML models require large, well-structured datasets for training—something not all organizations possess. Furthermore, the lack of transparency in complex models such as neural networks hinders decision-maker adoption (Barbierato & Gatti, 2024).

Application to Weapon System Development

Weapon system development projects are characterized by long durations, high uncertainty, and tight integration of subsystems—making them particularly susceptible to cascading delays. PDA offers an opportunity to shift from reactive project recovery to proactive risk management.

ML models can detect early signals of schedule slippage by integrating real-time data from testing cycles, supplier schedules, and subsystem integration reports. This approach mirrors that used by Awada et al. (2021), where field data integration enhanced mid-project forecasting (Awada et al., 2021).

Furthermore, hybrid approaches can enable defense agencies to tailor models using historical program data and expert judgment (Fitzsimmons et al., 2022). This would allow nuanced modeling of risk factors such as integration complexity, geopolitical disruptions, and budgetary constraints. PDA and ML can transform schedule management in weapon system development—enabling agile responses to risk, optimizing resource allocation, and increasing the likelihood of on-time, on-budget delivery of critical defense capabilities.

Acquisition Data

DTs, PDA, and ML require foundational historical data. The Department of Defense (DoD) collects data throughout the development process, including databases maintained by OSD and CADE. The intense DoD focus on cost dictates the data collected and its format. At the risk of adding more work, we suggest a review to collect and format data that can support schedule-focused PDA. ML could better predict schedule problems and address cost issues resulting from schedule delays.

As noted above, delay analysis requires that delay data be integrated into the model or the DT. Scheduling necessitates an analysis of the factors that have historically contributed to prolonged development times. Over the past 20 years, extensive research has identified several contributing elements to schedule delays, (Drezner & Smith, 1990; Van Atta et al., 2015). These factors include budget constraints, funding issues, complexity, technical challenges, and requirements. Building on these studies, research performed in 2018 used Selected Acquisition

Reports (SAR) from the OSD acquisition databases to identify delay factors cited by project managers during their annual SAR submissions (Pickar, 2018). Table 3 is a list of schedule delay factors developed in that study.

Table 3. Schedule Delay Factors
(Pickar, 2018)

Schedule Delay Factor
Administrative changes to schedule include updates to the APB and ADM as well as changes resulting from Nunn-McCurdy processes and program restructuring
Technical
Testing delays
Delay in the availability of key capabilities/ facilities (launch vehicle/ testing facilities/ IOT&E units)
Budget/ Funding Delays
Delays attributed to the Contractor
Delays because of Rework
External events such as inflation, earthquakes, labor strikes, etc. *(Force Majeure)*
Delays due to Contracting/ Contract Negotiation/ Award delays

Figure 2 is the proposed DT planning and execution model, with the planning aspect highlighted in red. The DT model is initially used for planning as part of the planning process. The work breakdown structure and task duration estimates are included in the input data (shown as data sources). The initial schedule data is the raw CPM-derived schedule, which is the basis for inputting data into the model. A key part of identifying and preparing the acquisition data is filtering by type of system (aircraft, missile, etc.). Schedule Delay Factors with frequencies are compared to the WBS and other planning data to determine similar occurrences. Once the project plan is built, the AI enhancements provide the necessary data-driven adjustments and a risk estimate. Outputs are the completed schedule and initial identification of schedule risk. Other risk areas are also identified.

Figure 2. Planning Aspects of the DT Approach

Figure 3 is the complete model incorporating the planning aspects. The diagram illustrates a comprehensive framework for integrating the DT model into the weapon system development project management processes, focusing on enhancing schedule optimization and decision-making in complex projects. It outlines the dynamic interaction between data acquisition, core processes, AI enhancements, and decision support mechanisms.

Figure 3. DT Planning and Execution Model

The process starts with acquisition data, which feeds into the input data stage. As noted above, the acquisition data currently used for this study is from the DoD SAR. As data availability on active programs increases throughout the DoD, the available data will increase exponentially, improving the process. This stage consolidates various types of project-related data, including project schedule data, schedule delay factors, and other relevant data sources. The DT then utilizes these inputs, a virtual representation of the physical project environment that enables simulation and analysis in real-time.

The core processes, comprising planning and execution activities, directly interact with the DT. This interaction ensures that the simulation model remains synchronized with actual project developments and strategic planning efforts. The DT generates key outputs such as an optimized schedule, risk mitigation strategies, and resource forecasting. These outputs are critical for effective project management and feed into the decision support system. This system offers dashboard insights, automated alerts, and proactive recommendations to support informed and timely decision-making by project stakeholders. An essential part of the model are the feedback loops throughout. As information is received, the model can be continuously updated to improve the execution.

Enhancing this entire framework are the AI enhancements, which integrate advanced analytical capabilities, including PDA, AI-driven schedule adjustments, and resource allocation optimization. These AI modules provide feedback to the DT, allowing for iterative improvements and adaptive project planning.

The diagram encapsulates a closed-loop, intelligent project management ecosystem where a DT is the central analytical engine. It continuously ingests data, interacts with planning processes, and outputs actionable insights, all enhanced by AI-driven modules to optimize performance and reduce project risks.

Concluding Thoughts on Digital Twins

Integrating DT technology into project management systems allows organizations to monitor, simulate, and optimize projects in real-time. It enhances risk management, decision-making, collaboration, and efficiency, making project execution more predictable and effective.

Adopting DTs in project management changes how projects are planned, executed, and monitored. By offering real-time insights, predictive capabilities, and enhanced collaboration, DTs can address traditional project management challenges.

Future research should focus on standardizing DT implementations and exploring their integration with emerging technologies like blockchain and quantum computing.

Prediction Markets: Synthesizing Scattered Information

Introduction

“Executives know . . . valuable information is scattered across the organization. They just don’t know how to retrieve it” (Thompson, 2012, p. 1).

“Those who made that (Challenger launch) decision were unaware of the recent history of problems concerning the O-rings” (Rogers Report, 1986, p. 88).

Our joint line of inquiry is focused on using relevant information to manage acquisition programs more effectively. However, data-driven actions are unlikely to be more efficacious than quality of the data, as noted above. Effective program management is more likely to be achieved through information derived from various sources and methods. In this part of our paper, we consider information derived from the program team members—suppliers and DoD program managers. As experience shows (e.g., the Challenger mishap), failure to make informed decisions (in acquisition and operations) is much easier said than done.

Along these lines, much information resides inside an organization (Thompson, 2012), but obtaining high-quality information through regular channels has often proved difficult. But well-organized markets have shown potential for eliciting information (Hayek, 1945, pp. 17, 19–23). In this context, prediction markets are a relatively recent method to elicit useful information—potentially valuable for defense acquisition managers.

The following sections focus on actual cases of markets’ (including prediction markets’) potential for aggregating information useful for acquisition program decisions. We also identify some potential problems for prediction markets in a defense acquisition context.

Identifying the Challenger Accident's Causes

The stock market's response to the Challenger Space Shuttle loss on January 28, 1986 demonstrated a well-organized market's ability to gather and sift information. Challenger launched its 10th mission. Shortly after getting airborne, the Shuttle experienced a catastrophic failure of its booster rockets. The Challenger was destroyed, with the loss of all crew members.

President Reagan directed a major investigation conducted by a select commission—chaired by former Attorney General William Rogers. The Commission was formed on February 6, issued its report on June 6 (Rogers Commission, 1986, pp. i, iii), and found that failure of O-rings intended to keep rocket engine thrust properly contained was the sole (proximate) cause of the accident. The report (Rogers Commission, 1986, p. 45) stated this conclusion:

> the loss of the Space Shuttle Challenger was caused by a failure in the joint between the two lower segments of the right Solid Rocket Motor. The specific failure was the destruction of the seals that are intended to prevent hot gases from leaking through the joint during the propellant burn of the rocket motor.

By way of root causes, the Rogers Commission (1986, p. 88) noted the following. The launch decision "was flawed. . . . Those who made that decision were unaware of the recent history of problems concerning the O-rings . . . *If the decision-makers had known all of the facts*, it is highly unlikely that they would have decided to launch" (Rogers Commission, p. 88, emphasis added).

While the Rogers Commission conducted its investigations and deliberations, stock market investors incorporated information on the Challenger's loss—concerned, *inter alia*, with the adverse effects on the company's stock price, whose product was the proximate cause of that loss. As Maloney and Mulherin (2003, p. 453) reported, "in the period immediately following the crash, securities trading in the four main shuttle contractors singled out the proximate cause of the accident (indirectly) by identifying the firm that manufactured the faulty component."

We note two significant features of this story. First, while the infamous O-rings received more publicity, the failure to make a properly informed launch decision was more consequential. Second, the Rogers Commision (1986, p. 88) makes clear that risks associated with colder-weather launches was available but somehow did not find its way to those who actually chose to launch on January 28. This is impressive support for Thompson's (2012) assertion cited above.

However, there is fairly strong evidence that interested participants in the stock market quickly focused on those O-rings and behaved accordingly. In fact, there was a suspension in trading of Thiokol[1] shares for a period of time (Maloney & Mulherin, 2003, p. 453). In short, the marketplace, on this occasion, performed according to Hayek's contention that markets are highly effective processors and aggregators of decentralized information.

Could a well-designed prediction market have gotten the relevant facts to those decision-makers? Observable events strongly suggest that a stock market could sort out the causes of the Challenger disaster—albeit *ex post*.

The 2024 Presidential Election: Prediction Markets vs. the Pollsters

During that election, the standard polls and "experts" predicted a very close race (270 to Win, 2024). One forecasting exercise gave Harris a 50.12% chance of winning (based on an extensive simulation exercise; Osipovich, 2024a). As it turned out, Trump won handily: electoral college, popular vote, and the "blue wall" states. He won all his "safe" states and also won many "contested" states (270 to Win, 2024; NBC Chicago, 2024).

[1] The manufacturer of those O-rings.

In this case, the prediction markets did better than the pollsters (Ferguson & Rincon-Cruz, 2024). One major participant (the "Trump Whale") turned out to be prescient: picking a Trump win in the electoral college, the popular vote nationwide, and the primary battleground states (Osipovich, 2024c).

Inquiries as to why this happened led to multiple hypotheses. One was a right-wing conspiracy—intended to create a Trump bandwagon effect and discourage potential Democratic voters (Osipovich, 2024c). That the Trump Whale turned out to be a non-U.S. citizen who claimed to be interested in making money (Osipovich, 2024c) cast doubt on that particular hypothesis.

A second hypothesis was that prediction markets were better suited for forecasting this election (and probably others). In our opinion, the rationale centers on incentives:

- Those intending to vote for Trump were said to be shy about admitting it (Osipovich, 2024c)—possibly due to his widespread vilification (e.g., Fitton, 2025) before and throughout the campaign.
- Getting a representative sample through traditional polling methods has become more difficult for several reasons. Those inclined to participate in a survey likely reflect a sort of selection bias.[2] That is, willingness to participate in a polling survey was correlated with their propensity to vote for a particular candidate. In this context, the shy-Trump-voter hypothesis could well have significant explanatory power (Osipovich, 2024c).
- The most serious allegation against the polling establishment was tailoring its products to the interests of major news sources. The rationale is that too-close-to-call polling reports are more newsworthy, and therefore, the business interests of the polling agencies were more closely aligned with the too-close-to-call reporting customer interests than identifying the Trump advantage (Osipovich, 2024c).

There are good reasons to believe this particular case reveals a great deal.

First, the prediction markets were indeed better at estimating election results than the regular polls. This is consistent with the hypothesis that prediction markets are often better than "experts" in generating information (Ferguson & Rincon-Cruz, 2024).

Second, incentives matter, and prediction markets provide (at least) pretty good incentives to be right. And there is a good reason that prediction markets work well even with relatively small stakes (Servan-Schreiber, 2004; Yeh, 2006).

Carmageddon

From July 15 to 18, 2011, a stretch of Interstate 405 in the Los Angeles area was closed for improvements. The associated traffic problem was confidently expected to be severe. As one observer put it, "I feel our collective psyches might not be able to withstand a traffic jam of this magnitude" (Gostar, 2011). In the event, however, traffic in the area in question was significantly lighter than normal, and the massive traffic jam simply didn't happen.

Why was this so? A good general answer is self-negating prediction. More specifically, the closure was announced well in advance and motorists were well informed about construction schedules and alternative routes. It appeared that the reason for the unexpectedly light traffic was motorists' behavior—likely based on trust in the consensus prediction.

- They stayed home as advised (Gish, 2011); more likely, they rescheduled trips so as to be at home during the expected congestion.

[2] In part due to widespread ability to screen incoming telephone calls.

- Information about alternate routes kept traffic away from the construction area.
- The availability of public transportation, such as the Metrolink commuter rail system, kept some potential motorists off the highways.[3]

In short, motorists believed the traffic congestion forecasts and took steps to avoid the situation. As a result, traffic difficulties were much less than the consensus prediction.

If there had been a prediction market about the size of the traffic snarl, it's possible the market outcome would have been a high probability of significant congestion. It's also likely that market participants would have gotten wind of motorists' plans and reflected the new information in their bets.[4]

The "Circularity" Issue

Along these lines, the "Trump Whale" in the 2024 election prediction market activity was a most interesting development. "Theo" (a pseudonym) placed several large bets on Trump's success in multiple markets using multiple account identifiers (Osipovich, 2024a).

Initially, there was some concern from the establishments (both polling and political) that Theo intended to create a synthetic groundswell of support for Trump's candidacy and engender a band-wagon effect that would improve his chances of being elected. "Theo," however, stated that he was in the prediction markets simply to make money—as discussed above (Osipovich, 2024b). Among other things, he commissioned polls whose core question was how the subjects' neighbors would vote—hypothesizing that Trump voters were reluctant to reveal their intentions, given the size and ferocity of the anti-Trump movement. Theo reported that this approach yielded more accurate information than standard polling methods.[5]

What seems particularly important, however, is that high-stakes prediction markets can create perverse incentives. Suppose, for example, that some "whale" places bets in a high-stakes terrorism prediction market—and is interested in making lots of money. Given the stakes, it is conceivably possible for someone to commission an act of terrorism in keeping with his position in the market. If a "whale" could contract with anonymous pollsters for information on voters' intentions, then contracting with a terrorist organization for an event that made his bet correct could be feasible—with significant financial returns for both parties. It's also conceivable that an agent of a terrorist organization could participate in a prediction market—seeking to make a terrorist event also result in financial benefits.

There is reason to conclude that market results affecting real-world outcomes is a major concern. For example, prediction market results indicated a significant increase in the Bush Administration's election chances in 2004 if Osama Bin Laden were killed or captured prior to November of that year (Thompson, 2012, p. 53)[6]. There is some possibility of prediction market results influencing the actions of decision-makers using market results to influence events.

This gets us to the more general issue of "circularity."[7] As generally organized, prediction markets generally assume (even if tacitly) that the event under consideration is not affected by the market's operation. However, the state of the market itself (generally as the probability of Event A) is open to inspection—to attract potential betters, if nothing else.

[3] For example, the Metrolink (a commuter rail system) set records for ridership during the construction period (The Source, 2011).
[4] Sometimes called "Bayesian Updating."
[5] Since Theo has, at this time, not disclosed the details of the polling he commissioned, one cannot verify the polls' results.
[6] There is, however, we don't know of any indication that market outcome influenced counter-terrorism operations in 2004.
[7] "Self-fulfilling" and "self-negating" beliefs are similar concepts. These are part of standard economic discourse.

This raises some interesting questions. At a minimum, the market provides information to decision-makers who have some control over Event A, such as when or if the event occurs. If the market results are deemed useful, affected decision-makers will likely include the prediction market's equilibrium probability in their calculations. This has been called the "circularity" problem (Thompson, 2012, pp. 145–6). Actors in the situation (inside or outside of the market) could take actions to affect events to further their own interests.

Suppose further that the prediction market offers a (equilibrium) probability for Event A. Suppose also that probability is conditioned on some other event, B. Suppose finally that agents who could control the probability of Event B are aware of the market results. They now have an incentive to take steps to increase (or decrease) the likelihood of B. One prediction market reported a substantial increase for President Bush's 2004 reelection if Osama bin Laden were captured. As noted above, this could have provided a significant incentive to increase efforts to apprehend bin Laden (Thompson, 2012, p. 53).[8]

More broadly, it's well known that beliefs about an event can affect the likelihood of the event occurring. The closure of a stretch of Interstate 405 in the Los Angeles area was widely publicized in advance as an extreme case of traffic congestion. In the event, however, traffic was lighter than usual, with transit times also less than normal. One reason cited was the shift to public rail transportation (The Source, 2011). The Carmageddon forecast indicates timely information could cause management to take actions that would avoid untoward developments (or mitigate them). That is, self-negating predictions are highly desired results in prediction markets in a program management context. If so, how do the market authorities determine the winners? And if they can't, then how does the market incentivize the participants?

Perhaps cleverly posed questions could solve (or lessen) this problem. If so, what would be the source for those questions, and by what means would they be "admitted" to the prediction marketplace?[9]

However, another question is whether a prediction market could provide helpful, timely information regarding launch temperatures' effects on the O-rings. Even if the proper questions were proposed before the fatal Challenger launch, could a prediction market have responded to provide timely information?

Concluding Thoughts on Information Gathering

For standard prediction markets, the circularity issue is a potential problem. Circularity can be a primary goal for prediction markets devoted to defense acquisition programs. If the current policy is highly likely to cause a significant, untoward result, then program management can perhaps take action to avoid or mitigate that result. The problem then becomes sorting out who placed the right bet. If not addressed, such possibilities substantially reduce the propensity for serious participation in the prediction markets. This seems a thorny issue, with some potential for entering a "wilderness of mirrors."[10]

While it's possible that well-designed betting propositions can avoid (or lessen) this problem, doing that would likely take some creative thinking.

More fundamentally, there are remaining questions regarding the potential of prediction markets to better inform program management decisions. This suggests further inquiries, including those we identify here. Since well-organized markets are generally more effective:

---

[8] However, we know of no indication that the Administration was influenced by this result.

[9] For example, PredictIt, a political prediction market site, had 13 U.S. markets in session on March 19, 2025. No indications of how to open a market.

[10] A popular (and descriptive) term in the intelligence literature—in both its factual (e.g., Martin, 1980) and fictional manifestations.

getting a better focus on the rules and customs of prediction market operations—to include, perhaps, field interviews; and conducting “gaming” studies of acquisition-oriented prediction markets in an experimental setting.

References

270 to Win. (2024). *Compare the 2024 presidential consensus and polling maps*. https://www.270towin.com/

Arditi, D., & Pattanakitchamroon, T. (2006). Selecting a delay analysis method in resolving construction claims. *International Journal of project management*, *24*(2), 145–155. https://www.sciencedirect.com/science/article/pii/S0263786305000906

Attaran, M., & Celik, B. G. (2023). Digital twin: Benefits, use cases, challenges, and opportunities. *Decision Analytics Journal*, *6*, 100165. https://doi.org/10.1016/j.dajour.2023.100165

Awada, M., Srour, F. J., & Srour, I. M. (2021). Data-driven machine learning approach to integrate field submittals in project scheduling. *Journal of Management in Engineering*, *37*(1), 42. https://doi.org/10.1061/(asce)me.1943-5479.0000873

Barbierato, E., & Gatti, A. (2024). The challenges of machine learning: A critical review. *Electronics*, *13*(2), 416. https://doi.org/10.3390/electronics13020416

Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5–32. https://doi.org/10.1023/a:1010933404324

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*, 273–297. https://doi.org/10.1007/Bf00994018

Drezner, J. A., & Smith, G. K. (1990). *An analysis of weapon system acquisition schedules.* RAND. http://scholar.google.com/scholar?q=related:Su8l_4wFOQ8J:scholar.google.com/&hl=en&num=20&as_sdt=0,5

Ferguson, N. & Cruz, M. R. (2024, November 16). How the Trump whale and prediction markets beat the pollsters in 2024. *Wall Street Journal*. https://www.wsj.com/politics/elections/how-the-trump-whale-and-prediction-markets-beat-the-pollsters-in-2024-dd11ec4e

Fitton, T. (2025, January 3). President Trump needs to clean house. *The Verdict*.

Fitzsimmons, J. P., Lu, R., Hong, Y., & Brilakis, I. (2022). Construction schedule risk analysis—A hybrid machine learning approach. *Journal of Information Technology in Construction*, *27*, 70–93. https://doi.org/10.36680/j.itcon.2022.004

Franck, R., & Pickar, C. (2024). *Navigating the labyrinth: Unraveling schedule complexity*. https://dair.nps.edu/handle/123456789/5114

Ghimire, S., & Mishra, A. K. (2019). Comparative study of prospective delay analysis techniques (DATs). *Saudi Journal of Civil Engineering*, *03*(05), 84–98. https://doi.org/10.36348/sjce.2019.v03i05.001

Gish, J. (2011). *I-405 summer weekend closure planned*. Inside Seven (CALTRANS). https://web.archive.org/web/20150912102517/http:/www.dot.ca.gov/dist07/Publications/Inside7/story.php?id=651

Gondia, A., Siam, A., El-Dakhakhni, W., & Nassar, A. H. (2020). Machine learning algorithms for construction projects delay risk prediction. *Journal of Construction Engineering and Management*, *146*(1), 04019085. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001736

Gostar, R. & Burreson, S. (2011). Carpocalypse: The weekend the 405 freeway will stand still. *Brentwood Patch*. https://web.archive.org/web/20120718034543/http:/brentwood.patch.com/articles/carpocalypse-the-weekend-the-405-freeway-will-stand-still

Grieves, M. (2002). *Completing the cycle: Using PLM information in the sales and service functions [slides]*. In *SME Management Forum*.

Grieves, M., & Vickers, J. (2017). Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In *Transdisciplinary Perspectives on Complex Systems* (pp. 85–113). Springer International Publishing. https://doi.org/10.1007/978-3-319-38756-7_4

Hayek, F. A. (1945). American economic review, (vol 35), 17–31. (Reprinted from *Price theory: Penguin modern economics readings,* by H. Townsend, Ed., 1971.).

Kerzner, H. R. (2022). *Project management: A systems approach to planning, scheduling, and controlling* (13th ed.). John Wiley & Sons.

Lee, B. (2017). *Proactive project management*. University of Chicago Press.

Liu, M., Fang, S., Dong, H., & Xu, C. (2021). Review of digital twin about concepts, technologies, and industrial applications. *Journal of Manufacturing Systems*, *58*, 346–361. https://doi.org/10.1016/j.jmsy.2020.06.017

Loh, W. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 14–23. https://doi.org/10.1002/widm.8

Martin, D. C. (1980). *Wilderness of mirrors*. Ballantine Books.

Marzouk, M. M., & El-Rasas, T. I. (2014). Analyzing delay causes in Egyptian construction projects. *J Adv Res*, *5*(1), 49–55. https://doi.org/10.1016/j.jare.2012.11.005

NBC Chicago. (2024, November 5) What are the “blue wall” states?. https://www.nbcchicago.com/news/local/what-are-the-blue-wall-states-what-to-know-as-results-come-in/3593443/

Osipovich, A. (2024a, October 18). A mystery $30 million wave of pro-Trump bets has moved a popular prediction market. *Wall Street Journal*. https://www.wsj.com/finance/betting-election-pro-trump-ad74aa71

Osipovich, A. (2024b, November 1). The Trump-betting whale speaks out: “I have absolutely no political agenda”: Self-described French trader says he bet more than $30 million US election after analyzing polls.

Osipovich, A. (2024c, November 6). How the Trump whale correctly called the election. *Wall Street Journal*. https://www.wsj.com/finance/how-the-trump-whale-correctly-called-the-election-cb7eef1d

Pan, Y., & Limao, Z. (2021). A BIM-data mining integrated digital twin framework for advanced project management. *Automation in Construction*, *124*, 103564. https://doi.org/10.1016/j.autcon.2021.103564

Pickar, C. K. (2018). *Informing DoD program planning through the examination of the causes of delays in acquisition using acquisition data*.

https://calhoun.nps.edu/bitstream/handle/10945/58762/SYM-AM-18-079-019_Pickar.pdf?sequence=1

PredictIt. (2025). https://www.predictit.org/markets

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, *1*, 81–106. https://doi.org/10.1007/BF00116251

Rogers Commission. (1986). *Report to the president on the space shuttle Challenger accident*. https://sma.nasa.gov/SignificantIncidents/ass

Schölkopf, B., Smola, A., & Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, *10*(5), 1299–1319. https://doi.org/10.1162/089976698300017467

Servan-Schreiber, E. et al. (2004, September). Prediction markets: Does money matter?. *Electronic Markets 14*(3), 9–10.

Silver, N. (2024, October 7). *What's behind Trump's surge in prediction markets?.* https://www.natesilver.net/p/whats-behind-trumps-surge-in-prediction

Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. (2019). Digital twin in industry: State-of-the-art. *IEEE Transactions on Industrial Informatics*, *15*(4), 2405–2415. https://doi.org/10.1109/tii.2018.2873186\

The Source (2011), *"Metrolink sets weekend ridership recording during 405 closure".*

Thompson, D. N. (2012). *Oracles: How prediction markets turn employees into visionaries*. Harvard Business Review Press, p. 1.

Uhlenkamp, J.-F., Karl, H., Wellsandt, S., Thoben, K.-D. (2019). *Digital twin applications: A first systemization of their dimensions*. In (p. 48). IEEE. http://dx.doi.org/10.1109/ice.2019.8792579

Van Atta, R. H., Kneece Jr, R., Patterson, C. M., & Hermes, A. C. (2015). Assessing weapon system acquisition cycle times: Setting program schedules. *Institute for Defense Analysis IDA*. http://www.dtic.mil/docs/citations/AD1005582

Yeh, P. (2006). Using prediction markets to enhance US intelligence capabilities: A "Standard & Poors 500 index" for intelligence. *Studies in Intelligence* (50/4).

Tactical Overmatch by Design: Acquisition Engineering for Smart Warfare

Louis Libin—holds an MS in Electrical Engineering with a concentration in Physics from the Belfer Graduate School of Science and a BS in Electrical Engineering from Pacific Western University. Libin completed an optical communications graduate program at MIT. He completed the Executive Leadership Program at the Center for Homeland Defense and Security, at the U.S. Naval Postgraduate School. Libin has completed almost 30 military and defense-related course including the U.S. Army War College, USAWC Commandants, National Security Program.

Libin served as Chief Technology Officer at General Electric for 13 years and has since founded four technology-driven companies. He is the founder of Broad Comm, a firm specializing in international wireless development, defense spectrum management (public safety), and operational communications across global conflict zones and high-demand civilian environments.

Internationally recognized for his leadership in wireless and spectrum policy, Libin serves as a Special Rapporteur to the International Telecommunication Union (ITU) in Geneva on wireless standards. Domestically, he serves on the U.S. Secret Service Spectrum Committee and chairs the Wireless Coordination Committee for major U.S. political events, including national political conventions, Election Day operations, and Presidential Inaugurations.

Libin has designed and instructed numerous academic and professional development courses focused on the intersection of military strategy, spectrum theory, and emerging technologies. His teaching integrates core concepts of physics, wireless engineering, and asymmetric warfare, with a focus on service learning, team-based operational planning, and experiential analysis.

He has deployed to active conflict zones—including Ukraine and Gaza—in support of U.S. military missions, where he provided advisory expertise on military innovation, conducted field-based operational analysis, and documented the impact of emerging technologies on modern warfare for strategic assessment. [louislibin@broad-comm.com]

Abstract

The modern battlefield is evolving under the influence of rapid technological advances in wireless communications and drone systems. These systems are no longer just support tools—they form the core of tactical superiority. This paper proposes a unified framework for achieving tactical overmatch through deliberate acquisition engineering. Drawing on battlefield lessons from Ukraine and Israel-Gaza, it details the convergence of software-defined networks, loitering munitions, and autonomous targeting systems. The argument is built around the thesis that acquisition reform must be engineered for speed, modularity, and interoperability. Recommendations center on digital twin validation, AI assurance protocols, zero-based budgeting, and dynamic field feedback loops. Only by redesigning the acquisition process to reflect the speed of modern warfare can the Department of Defense guarantee dominance in future conflicts.

Keywords: Tactical dominance, wireless communications, UAS, loitering munitions, acquisition reform, interoperability, modular design, digital twin, cyber resilience, AI assurance

Introduction

The dynamic evolution of wireless communications and unmanned systems is reshaping modern combat. From Ukraine's drone swarms to Israel's seamless battlefield coordination, today's warfare demands smarter tools and faster integration. However, outdated acquisition practices inhibit rapid fielding of these technologies. This paper argues that the U.S. Department of Defense (DoD) must adopt an acquisition engineering model that integrates modularity, cyber resilience, and real-time field feedback. Tactical overmatch is not accidental, it is engineered.

The Tactical Value of Wireless Communications and Unmanned Systems

Ukraine and Israel-Gaza as Case Studies

Ukraine's use of decentralized drone swarms and Elon Musk's Starlink network has shown how flexible communication networks and low-cost drones can reshape a battlefield. Meanwhile, Israel has perfected the synchronization of loitering munitions and wireless targeting in dense urban environments, as seen during the 2023 Gaza conflict.

Converging Technologies

Tactical superiority now depends on integrating Loitering Munitions (LM), Unmanned Aerial Systems (UAS), and Tactical Mesh Networks. These systems generate terabytes of data, requiring robust and adaptive wireless architectures. The battlefield is becoming a live data ecosystem.

Table 1. Comparative Assessment of Tactical Tech Usage

Region	Technology Focus	Tactical Outcome
Ukraine	Starlink, DJI drones	Disrupted Russian command and logistics
Israel-Gaza	Loitering munitions, C4I	Rapid neutralization of high-value targets

Current Acquisition Gaps

Speed Deficiencies

The traditional Defense Acquisition System (DAS) cycle—Concept > Development > Testing > Procurement—takes 7–15 years. On today's battlefield, that's an eternity.

Cybersecurity Fragmentation

Platform-centric procurement often lacks integrated cybersecurity from the ground up. Systems are patched after deployment rather than designed for cyber resilience.

Interoperability Challenges

Vendors push proprietary interfaces. As a result, drones, radios, and artificial intelligence (AI) systems often fail to communicate across branches or with allies.

Acquisition Engineering for Tactical Overmatch

Digital Twin Environments

Digital twins allow real-time testing of equipment in simulated battlefield conditions. All systems should be tested against adversarial jamming, GPS spoofing, and denied environments.

Modular Open Systems Architecture

All acquisitions must meet modular open systems architecture (MOSA) standards. Interchangeable sensors, payloads, and control systems cut costs and speed up integration.

AI Assurance Protocols

Machine learning models used in targeting or threat analysis must undergo continuous adversarial testing. Explainability and bias detection are critical. The DoD should require formal AI red-teaming.

Cybersecurity by Design

Zero Trust Architecture and secure firmware updates must be embedded in design, not retrofitted post-fielding.

Human-System Integration Testing

New systems must undergo usability testing with real-world operators. Systems that overwhelm the user reduce battlefield efficiency.

Battlefield-Driven Procurement: A Feedback Loop Model

```
[Field Ops] → [Data Capture] → [Dev/Test in Sim] → [Rapid Prototype] →
[Deploy] → [Field Ops] → …
```

Figure 1. Smart Acquisition Feedback Loop

This loop minimizes bureaucracy while validating systems directly in operational environments.

Budgeting and Contract Reform

Zero-Based Budgeting

Zero-based budgeting (ZBB) ensures every program component is justified annually. No more “use it or lose it” spending.

Fast-Track for Mission-Critical Prototypes

Congress should authorize a permanent Other Transaction Authority (OTA) path for prototypes under $50 million linked to frontline utility.

Dual-Use Incentives

Companies building commercial 5G, IoT, or AI systems should receive DoD tax credits for modifying their products for military use.

Decentralized Procurement Cells

Empower units with their own procurement officers trained in acquisition engineering to address specific tactical needs.

Adaptive Contracting Models

Use outcome-based and rolling contracts that emphasize iterative deliveries and operational validation.

Table 2. Suggested Acquisition Metrics by Phase

Phase	Metric	Ideal Benchmark
Prototyping	Time to field	< 6 months
Integration	Interoperability Index	90%+ compatibility
Security	Penetration test pass rate	100% against known threats
Usability	Operator performance rating	90%+ satisfaction

Recommendations for Implementation

Create a Joint Tactical Engineering Office under the DoD to manage battlefield-tech acquisition

To streamline the integration and acquisition of advanced battlefield technologies, the DoD should establish a Joint Tactical Engineering Office (JTEO). This office would serve as a centralized hub to manage the development, testing, and deployment of battlefield technologies, ensuring that innovations from across the services are effectively synchronized. By centralizing expertise and oversight, the JTEO would reduce redundancy, enhance cross-branch collaboration, and ensure that all tactical systems meet the rigorous operational demands of the modern battlefield. This approach would facilitate faster decision-making, improve resource allocation, and ensure that new technologies are fielded in a timely manner.

Mandate MOSA compliance for all new tactical systems.

Mandating MOSA compliance for all new tactical systems would enable interoperability, flexibility, and scalability in the military's technology portfolio. By designing systems with open standards, the DoD can more easily integrate components from various manufacturers, ensuring that future upgrades and improvements can be made without replacing entire systems. This would foster innovation, reduce long-term costs, and allow for faster adaptation to evolving threats. MOSA compliance would also ensure that systems remain adaptable to future technologies, reducing the risk of obsolescence and enabling quicker responses to emerging battlefield needs.

Develop cross-branch digital twin simulation centers.

To enhance training and operational preparedness, the DoD should establish cross-branch digital twin simulation centers. These centers would use advanced simulation technologies to create virtual replicas of physical assets, systems, and battlefields, enabling real-time, data-driven analysis and testing of different scenarios. By allowing joint forces to simulate complex operations, cross-branch digital twin centers would foster interoperability, refine tactics, and optimize decision-making. They would also enable rapid testing of new technologies and systems before deployment, ensuring that innovations are field-tested in a virtual environment before they are introduced in the real world.

Adopt AI assurance protocols as part of milestone reviews.

As AI becomes an integral part of military operations, it is essential to incorporate AI assurance protocols into the DoD's acquisition milestone reviews. These protocols would ensure that AI-driven systems are rigorously tested for safety, reliability, and ethical compliance before being deployed. By embedding AI assurance into the acquisition process, the military can mitigate risks associated with autonomous systems and ensure that they function as intended in real-world conditions. This would build confidence in AI technologies while safeguarding against unintended consequences, ensuring that systems remain under human oversight and control.

Expand OTA use and ZBB-based budgeting models.

To accelerate the development and fielding of new technologies, the DoD should expand the use of OTAs and ZBB models. OTAs offer flexibility in acquiring innovative technologies by bypassing traditional acquisition processes, enabling faster collaboration with industry partners. Meanwhile, ZBB-based budgeting would require a fresh evaluation of each program's needs, ensuring that resources are allocated efficiently and aligned with the most urgent priorities. Together, OTAs and ZBB would enhance the DoD's ability to quickly adopt new technologies, adapt to changing priorities, and reduce waste in defense spending.

Require cyber threat modeling and mitigation prior to procurement.

Cybersecurity must be an integral part of the acquisition process for all new military systems. Requiring cyber threat modeling and mitigation strategies prior to procurement would ensure that potential vulnerabilities are identified and addressed before systems are deployed. This proactive approach would minimize the risk of cyberattacks, enhance the resilience of military technologies, and ensure that sensitive data remains protected. By incorporating cybersecurity considerations early in the acquisition process, the DoD can ensure that new systems are secure, reducing the likelihood of costly breaches and ensuring the integrity of critical military operations.

Incorporate usability testing and operator feedback into acquisition milestones.

The effectiveness of new tactical systems depends not only on their technical capabilities but also on their usability by the operators who rely on them in combat. Incorporating usability testing and operator feedback into acquisition milestones would ensure that systems are intuitive, user-friendly, and aligned with the needs of military personnel. Regular feedback loops throughout the development process would help identify and resolve operational challenges, improving system performance and reducing the likelihood of operational errors. This approach would prioritize the human element in technology design, ensuring that systems are not only advanced but also effective in the hands of soldiers.

H. Encourage industry-academic-military partnerships for tech transitions.

To facilitate the transition of cutting-edge technologies into military applications, the DoD should foster stronger partnerships between industry, academia, and the military. These collaborations would combine the innovative capacity of the private sector with the expertise of academic researchers and the operational experience of the military. By creating a more dynamic and collaborative ecosystem, the DoD can accelerate the development and deployment of next-generation technologies, while ensuring that these solutions meet the unique needs of military operations. These partnerships would also provide a continuous feedback loop that fosters ongoing innovation and ensures that military technologies remain at the forefront of global advancements.

Conclusion

The pathway to smart warfare lies not only in adopting new technologies but in engineering acquisition processes that foster tactical overmatch by design. Wireless communications and drone systems have proven their critical role in asymmetric warfare, yet they remain under-leveraged due to outdated acquisition models. With bold shifts—digital twins, modular open systems, cyber-secure platforms, and combat-ready prototyping—the DoD can transform procurement into a force multiplier. This engineering-first approach to acquisition ensures U.S. forces remain ahead of adversaries across every domain.

Disclaimer

The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

References

AFWERX. (2024). *Innovation in defense acquisition: Case studies and frameworks* (U.S. Air Force Publication No. 44-001). U.S. Air Force.

Defense News. (2023). Starlink in Ukraine: A game changer in battlefield communications. https://www.defensenews.com

DoD. (2023). *Defense acquisition system modernization: Strategy and reform initiatives*. Office of the Under Secretary for Acquisition and Sustainment.

Israeli Defense Forces. (2023). *Unit 8200 cyber defense capabilities briefing*. IDF Cyber Directorate Internal Report.

Israeli Defense Review. (2023). Iron dome: Evolution and applications in counter-UAS defense. 12.

Jane's Defense Weekly. (2022). Bayraktar TB2: A tactical analysis of combat effectiveness. 59(3).

L3Harris Technologies. (2022). *AN/PRC-117G product specification sheet*. https://www.l3harris.com

NATO Communications and Information Agency. (2023). *Federated mission networking implementation guide*.

RAND Corporation. (2021). *Counter-UAS technologies: Trends and developments in the tactical space* (Report No. RR-A789-1).

U.S. Army. (2022). *Warfighter Information Network-Tactical (WIN-T) program overview*. Army Network Modernization Strategy, 11–18.

Simplifying the Complex: A Conversational Approach to Configuring Military Simulators

Protima Banerjee—has held technical and leadership positions designing and developing high performance data and computing systems in the DoD and commercial sectors for almost 30 years. Her contributions are part of critical programs provided to the U.S. Navy, U.S. Coast Guard and associated foreign military sales. Banerjee is a Technology Fellow at ASRC Federal and an Adjunct Professor of Computer Science at Rowan University. Her research interests include conversational interfaces, improving AI-driven search answer validation through semantic methods, optimizing large-scale data processing, advancing human-AI collaboration, and developing resilient AI models for mission-critical applications. [pbanerjee@asrcfederal.com]

Abstract

Complex military software systems, such as simulators and decision support tools, often necessitate extensive user training to master configuration tasks. This research proposes a novel approach to streamline user interactions with these systems by leveraging the capabilities of large language models in conjunction with semantic information structured in knowledge graphs. By employing a conversational interface, inexperienced users can interact with complex systems using natural language, significantly reducing the learning curve and operational overhead.

Introduction

Military software systems, such as simulators and decision support tools, are integral to modern defense operations. These systems are often highly complex, involving many intricate configuration parameters that require significant training to setup correctly. Operators must possess a combination of deep domain knowledge and technical expertise; this type of skilled individual is difficult to attract, train and retain. Consequently, the learning curve associated with these systems can create operational bottlenecks, ultimately hindering mission agility and effectiveness.

This research introduces a novel approach to address these challenges by leveraging the capabilities of large language models (LLMs) in conjunction with semantic information represented in knowledge graphs (KGs). LLMs, known for their proficiency in natural language understanding, enable users to interact with complex systems using conversational language, drastically simplifying the configuration and operation of these systems. By pairing LLMs with the semantic and syntactic data representation capabilities of KGs, this approach offers a more intuitive and accessible interface. This conversational interface can enable domain experts to interact with simulation systems in a manner that feels natural and intuitive, substantially reducing the time and effort required for training while minimizing the potential for user error.

The remainder of this paper walks through our problem statement, background and related work in LLMs and KGs, describes our technical approach, experiment setup, and concludes with a discussion of our results and future work.

Problem Statement

Traditional user interfaces for military simulation systems are often based on menus, forms, and commands and are not always intuitive. These interfaces are designed with expert users in mind, but they may alienate or frustrate novice users and domain experts who are unfamiliar with the intricacies of the underlying system. As a result, simulation system operators often require significant training and a long onboarding process. This not only affects system performance but can also contribute to increased operational overhead.

In response to this challenge, there is an urgent need for more efficient and accessible interaction interface. In recent years, LLMs have emerged as a promising approach to user interaction for military systems (Mikhailov, 2023). This technology enables operators to communicate with complex software using everyday language and has been shown to be effective for tasks such as course of action generation. However, even with the advancements in modern LLMs, several key issues must be addressed before such conversational interfaces can be reliably used to interact with real-world military systems:

1. **Natural Language to Structured Data Translation:** One of the most critical challenges is translating human-readable natural language inputs into structured, machine-readable data. This involves accurately interpreting user queries, which can vary greatly in terms of phrasing, domain-specific terminology, and syntax, and converting them into a format that the underlying system can process—typically a structured data format like JSON. For complex military systems, the translation must not only be accurate but also capable of handling both general and domain-specific contexts, such as military terminology, operational constraints, and system requirements. Achieving this goal would make natural language interaction a viable alternative to traditional interfaces, reducing the cognitive load on users and eliminating the need for detailed system expertise.

2. **Data Integrity and Completeness:** Once a user query is translated into structured data, ensuring that the data is complete, consistent, and free from errors is critical. Inaccuracies or omissions in system configurations can have serious consequences, particularly in mission-critical environments. Prior to deployment, an interface must be able to identify and address common issues in user input, such as incomplete or contradictory instructions, and ensure that the generated data meets the necessary standards for use in military systems. This includes validating the integrity of the data against the system's operational rules and constraints, as well as providing feedback to the user when input conflicts or inconsistencies are detected.

This research explores the above topics and proposes a novel architecture that leverages LLMs in conjunction with semantic information in the form of a structured KGs.

Background and Related Work

LLM Use Cases

LLMs have emerged as a powerful tool for processing and analyzing vast amounts of information, improving decision-making, and enhancing human-machine interaction (Caballero, 2024). In this section we briefly describe a selection of the military-relevant use cases in which LLMs have shown promise.

Intelligence Analysis

One of the earliest domains to explore the use of LLMs in the military has been intelligence gathering and analysis as shown by Logan (2024) and Nitzl (2024). The volume of data generated from various sources—including satellite imagery, signals intelligence (SIGINT), open-source intelligence (OSINT), and classified reports—has outpaced traditional analytical methods. LLMs have been deployed to automate and enhance intelligence analysis by extracting key insights, summarizing reports, and identifying patterns or anomalies that may indicate threats.

Military Decision-Aids

LLMs have also been explored to aid military-decision making. This process is inherently complex, requiring commanders to synthesize information from multiple domains—land, air, sea, space, and cyber. LLMs improve the efficiency of current systems by helping military planners analyze battlefield conditions, generate courses of action (COAs), and evaluate mission risks (Goecks, 2024). These AI-driven systems can serve as an assistant to commanders by generating real-time reports, summarizing intelligence briefings, and suggesting potential responses based on historical data and current operational factors.

Additionally, LLMs usage for battlefield management systems is another area of active research (Connolly, 2024). These systems process sensor data, intelligence reports, and battlefield communications, allowing commanders to access critical information through natural language queries. By combining LLMs with knowledge graphs and structured data sources, military operators can retrieve highly relevant and contextual information without the need for extensive manual searching.

Cybersecurity Operations

The modern battlefield extends into cyberspace, where cyber warfare and digital threats pose significant challenges. LLMs have been increasingly employed in cybersecurity operations for automated offense and defense mechanisms (Anurin, 2024). In this domain, LLMs can be used to analyze vast amounts of cyber threat intelligence, detecting patterns of malicious activity and predicting potential vulnerabilities. Additionally, LLMs have been integrated into cybersecurity chatbots and virtual assistants to help analysts rapidly assess and respond to cyber incidents (Shafee, 2024).

Human Computer Teaming

Another promising domain area for LLM use in the military is human-computer teaming. The military has increasingly relied on autonomous systems, including unmanned aerial vehicles (UAVs), robotic ground units, and AI-driven mission control assistants. Effective communication between human operators and these autonomous systems is essential for mission success. LLMs have been explored as a means to enhance human-machine interaction by providing more intuitive and natural language interfaces (Javaid, 2024).

Logistics and Supply Chain Management

Efficient logistics and supply chain management are crucial for sustaining military operations. LLMs could be utilized to optimize logistics planning, streamline supply chain coordination, and predict equipment maintenance needs as shown in Aghaei (2025) and Olena (2024). By analyzing historical data and real-time logistical information, these models help military logisticians identify potential bottlenecks, improve inventory management, and ensure timely delivery of critical supplies.

One application of LLMs in logistics involves predictive maintenance (Lukens, 2023). AI-driven models analyze sensor data and maintenance records to forecast potential mechanical failures, allowing for proactive maintenance scheduling. This capability reduces downtime and enhances the overall readiness of military assets.

Regulatory Compliance.

Finally, LLM use has been explored for automation of compliance processes (Makovec, 2024). By reasoning through and automating some or all of the compliance process, LLMs have the potential to help reduce administrative workloads and improve overall efficiency in compliance operations.

Limitations of LLMs

Despite significant progress operationalizing LLM use, existing approaches face several limitations (Biswas, 2023):

- Contextual Understanding: LLMs struggle with domain-specific language understanding and operational contexts.
- Data Validation: Most systems lack robust mechanisms to identify and resolve inconsistencies or omissions in the generated data.

KGs

A KG is a structured representation of information that captures relationships between entities in a way that mimics human understanding (Hogan, 2021). Unlike traditional databases that store information in isolated tables, KGs use a network of interconnected nodes and edges to represent data as a web of relationships. This enables more intuitive data retrieval, contextual reasoning, and advanced analytics. KGs power applications like search engines, recommendation systems, and AI-driven assistants by enabling machines to understand and infer meaning from complex data (Peng, 2023). They are built using ontologies, making them particularly valuable for domains like the military and have the potential to play a crucial role in powering intelligent applications.

Technical Approach

Overview

The integration of LLMs and KGs offers a promising pathway to overcome the above limitations. By combining the natural language capabilities of LLMs with the semantic structuring power of KGs, it is possible to create natural language system interfaces that are both intuitive and reliable. These interfaces can facilitate the translation of user inputs into structured data, while ensuring data integrity through validation mechanisms. An overview of our technical approach and architecture is presented in Figure 1.

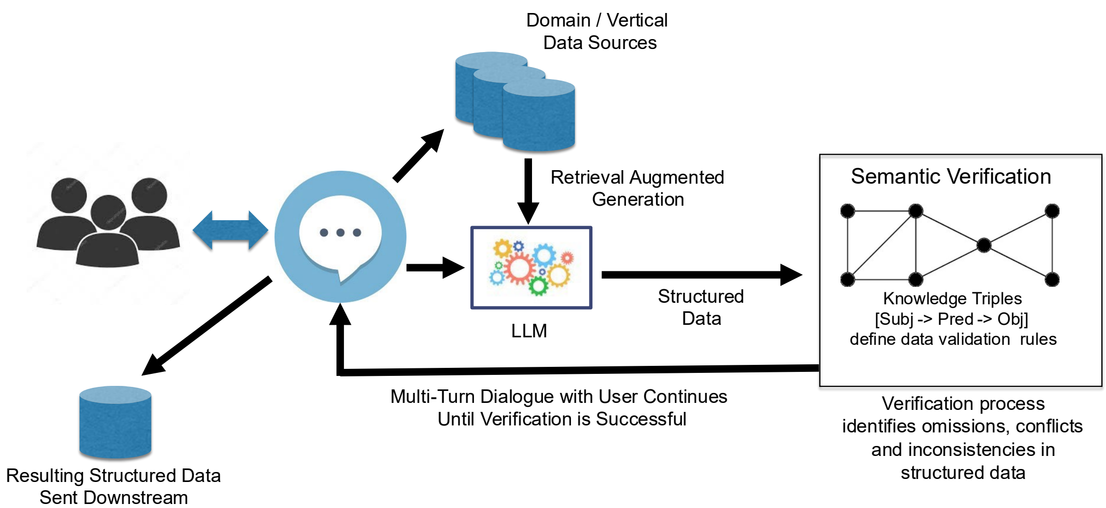


Figure 1. Technical Approach and System Architecture

The system design consists of several key layers working together to ensure accuracy, consistency, and usability of the final output. The process begins with a user interaction component presented as a chat interface. This interface manages communication and forwards user queries to the LLM for processing. The processing layer leverages retrieval-augmented generation (RAG; Gao, 2023), allowing the LLM to access relevant external knowledge sources, such as structured databases or KGs, to generate a well-formed structured data output.

Once generated, the structured data passes through to a semantic verification component, which checks for data inconsistencies, omissions, and conflicts. This layer applies a combination of rule-based logic and reasoning techniques to ensure the generated data aligns with known facts and domain constraints. If issues are detected, the system triggers a feedback loop, presenting the user with clear questions to direct the necessary adjustments. The user can then refine their input iteratively until all inconsistencies are resolved, ensuring high data integrity every for every single instance of generated data.

After validation, the finalized structured data is sent to the the military simulator to configure the simulation scenario. It is noted here that while we are focused on a miltary simulation system as the downstream data processing engine in this case, this architecture could be applicable to any system that requires complex structured data inputs.

Semantic Verification

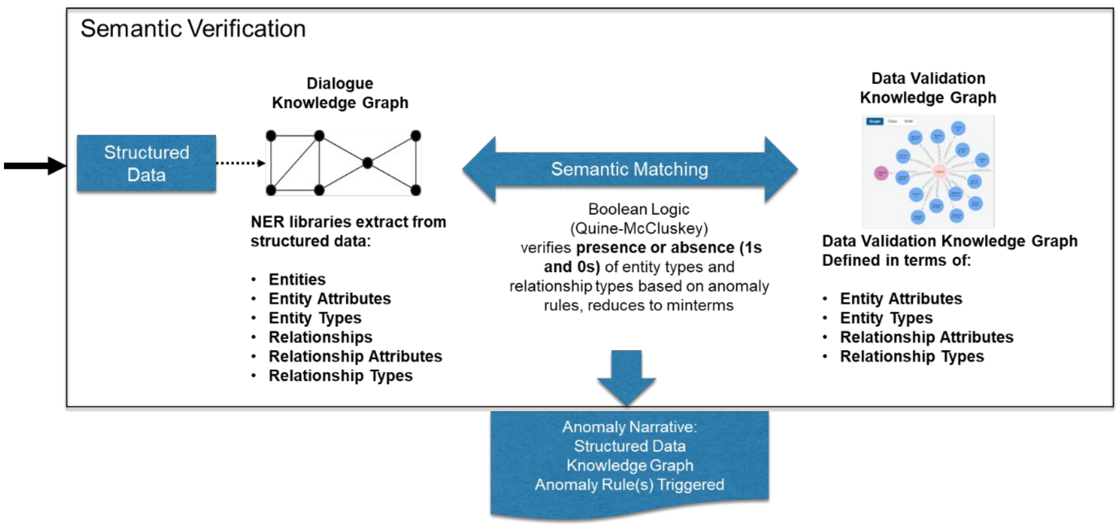


Figure 2. Design of the Semantic Verification Component

The semantic verification component shown in Figure 2 plays a critical role in ensuring the accuracy and consistency of structured data by leveraging knowledge graphs and Boolean logic algorithms. This component operates in multiple stages to validate structured data effectively.

First, the component constructs a KG from the structured data input. It utilizes Named Entity Recognition (NER) libraries to extract key elements, including entities, entity attributes, relationships, relationship attributes, and relationship types. This transformation ensures that the structured data is represented in a graph-based format suitable for comparison.

Next, the generated KG is compared against a predefined validation KG. This reference graph consists of known validation triples, defined in terms of entity attributes, entity types, relationship attributes, and relationship types. The goal of this step is to detect inconsistencies by examining the alignment between the structured data and the established validation rules.

To facilitate this matching process, the system employs Boolean logic algorithms, such as Quine-McCluskey, to systematically verify the presence or absence of required entity and relationship types. Each entity type and relationship type is mapped to a binary representation (1s and 0s), indicating whether they exist within the structured data. Using minimization techniques, the algorithm reduces these binary values to a minimal set of essential conditions (minterms) that highlight the key discrepancies.

Finally, the component generates an “anomaly narrative” for the user, detailing the specific validation rule(s) that were triggered due to inconsistencies. This narrative provides actionable insights, guiding the user to refine their input iteratively until the structured data fully aligns with the expected KG. By enforcing structured validation through formal logic, this approach ensures data integrity before passing the refined output to downstream systems.

One of the key advantages of this method is its ability to provide users with clear, actionable explanations of discrepancies. Instead of vague error messages, users receive precise feedback on which validation rules were triggered, allowing them to refine their input iteratively. This structured feedback loop ensures that only high-quality, validated data proceeds to downstream systems, reducing errors and improving decision-making.

The approach is also scalable, as KGs enable a structured representation of complex relationships, and the use of Boolean logic matching optimizes computational processing. By outputting only the minterms of the Boolean logic evaluation, the system reduces redundant validation checks, ensuring that the foundational inconsistencies are flagged. This efficiency is particularly beneficial in large-scale applications where structured data verification must be performed rapidly.

Technology Stack

Our implementation technology stack is shown in Figure 3.

	Category	Description
	Evaluation	Amazon Relational Data Store (RDS) for storing experiment results, Langchain Evaluation Framework
Knowledge Graph Toolset	Knowledge Graph Correlation	Cypher stored procedures running in Neo4j implement graph-matching algorithms, (Developed code)
	Graph Database	Neo4j (Containerized graph database, running on AWS compute)
	Knowledge Graph Generation	Langchain LLMGraphTransformer (Generates knowledge graphs from text, augmented with custom developed code for detection of complex relationships between entities)
LLM Toolset	LLM framework	Langchain (Langchain agents implement narrative management - addition and summarization)
	LLM (Hosted)	**gpt-3.5-turbo** accessed via OpenAI API, running remotely
	LLM (On-premise)	**Llama, Gemma, etc.** accessed via the Ollama framework, **running locally on AWS compute**
Infrastructure	Python	Python 3.10
	Hardware	Amazon Web Services (AWS) GovCloud (1 NVIDIA GPUs and vCPUs)

Figure 3. Implementation Technology Stack

Evaluation

Representative Dataset: Automated Identification System

Automated Identification System (AIS) data was selected as an ideal test case for our system prototype due to its structured yet moderately complex nature. This openly available maritime dataset (Kress, 2023) contains standardized information transmitted by commercial vessels and Coast Guard ships for collision avoidance.

AIS data is well-documented and includes essential mandatory fields such as ship identification number (MMSI), latitude, longitude, speed, and course. Additionally, several AIS fields require accuracy checks—for example, position and course must be correctly expressed in degrees, speed in knots, and timestamps in UTC. The need for consistency checks, such as ensuring that a ship's course and heading align and that its status corresponds logically with speed and heading, makes this dataset particularly relevant for detecting anomalies or inconsistencies. AIS domain knowledge is readily available from open sources such as NOAA AccessAIS (NOAA, 2024) and can be used to automatically or semi-automatically populate a domain knowledge rules as seen in Figure 4. The specific data set we used in these experiments consisted of 1000 AIS messages from vessel traffic around the port of New Orleans on March 31, 2022.

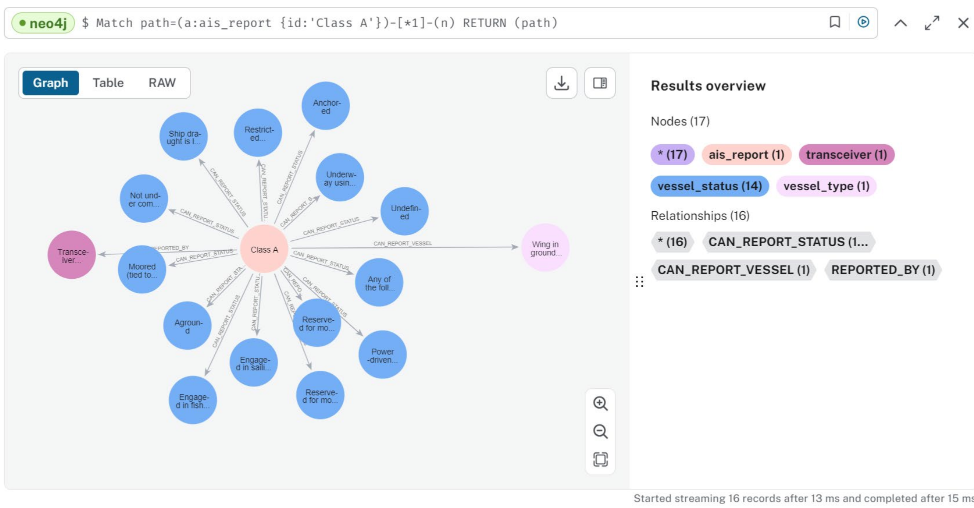


Figure 4. AIS Domain Rules Captured as a KG in Neo4J

Experiment Design

To evaluate the reliability of our technical approach, we designed a multi-step experimental framework, as shown in Figure 5.

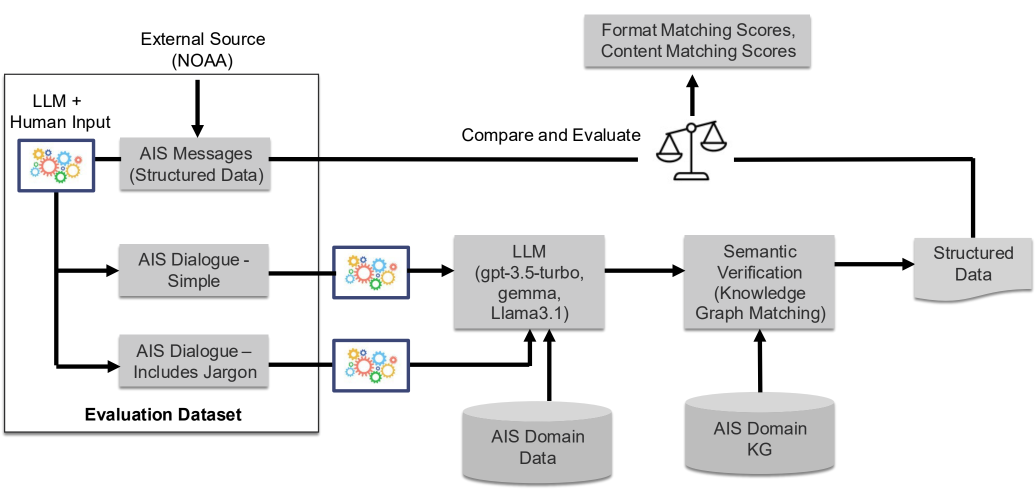


Figure 5. Experiment Framework and Setup

We began by establishing a source of truth dataset derived from structured AIS messages. We transcribed this structured data into natural language dialogue using a combination of LLM processing and human input. We created two distinct dialogue styles:

- Simple Dialogue—A straightforward transformation of AIS data into natural language, maintaining clarity and minimal linguistic variation.
- Colloquial and Jargon-Based Dialogue—A more complex transformation incorporating maritime jargon, conversational elements, and informal phrasing to mimic real-world human communication.

Next, we took the generated dialogues and processed them through various LLMs to reconstruct structured AIS data. We also incorporated our semantic verification process, ensuring that the extracted information adhered to expected AIS data structure formats and consistency rules. By varying the LLMs used in this process, we examined differences in their ability to infer structured data from both simple and jargon-heavy dialogue inputs.

Finally, we compared the reconstructed structured data to the original source of truth AIS dataset. The evaluation process generated format and content scores, assessing how accurately the LLM-driven reconstruction aligned with the original structured information. Format scores measured adherence to expected data structures (e.g., proper formatting in JSON, adherence to schema), while content scores quantified semantic accuracy, ensuring that key details such as course, heading, and ship status were correctly transcribed.

Our goal in this experiment was to determine how reliably our technical approach using LLMs in conjunction with a KG could convert human-generated dialogue into structured data. By analyzing performance across different dialogue styles and LLM models, we aimed to identify potential challenges and opportunities in using AI to extract structured data from human communication in the maritime data domains.

Experiment Results and Discussion

The results of our experiments are presented in Figure 6.

LLM	Dialogue Type	Semantic Verification	Format Score (%)	Content Score (%) Vessel Type Identification	Content Score (%) Status Identification
Gpt-3.5-turbo	Simple	No	100%	100%	94%
Gpt-3.5-turbo	Simple	Yes	100%	100%	99%
Gpt-3.5-turbo	**Jargon**	**No**	100%	24%	65%
Gpt-3.5-turbo	**Jargon**	**Yes**	100%	73%	79%
Llama3-8b	Simple	No	94%	100%	99%
Llama3-8b	Simple	Yes	95%	100%	99%
Llama3-8b	Jargon	No	94%	34.5%	40%
Llama3-8b	Jargon	Yes	95%	48%	38%
Gemma-7b	Simple	No	100%	100%	99%
Gemma-7b	Simple	Yes	100%	100%	99%
Gemma-7b	**Jargon**	**No**	100%	3%	16%
Gemma-7b	**Jargon**	**Yes**	100%	49%	39%

Figure 6. Experiment Results

Our experiments demonstrated that 100% format similarity, including strict adherence to JSON schema conformance, is easily achievable using state-of-the-art LLMs and LLM frameworks. Regardless of the complexity of the input dialogue, LLMs consistently produced structured outputs that matched the expected format.

However, content similarity presented more significant challenges. When the input dialogue was simple and direct, LLMs successfully extracted and translated the information into the correct structured data fields. In contrast, when jargon and colloquial language were introduced, accuracy dropped substantially. This was particularly evident in two key AIS fields: vessel type and vessel status. Both fields are encoded as numerical values in structured AIS data but would be described in natural language when reported by humans. For example, a vessel type code "31" could correspond to the words "Tug," "Tugboat," "Towing vessel," "ship assist vessel," etc. in dialogue, and an LLM must correctly map such descriptions back to their respective code values. Another example is status code "0," which could correspond to the words "underway," "at sea," "cruising," "sailing," etc. Accuracy in this mapping varied depending on the LLM, but we found that incorporating a KG significantly improved accuracy for two of the models tested (GPT-3.5-Turbo and Gemma:7B). The knowledge graph provided a structured reference, reducing ambiguity and improving the alignment between natural language descriptions and standardized AIS codes.

Several additional systematic errors were observed in the reconstructed structured data. LLMs exhibited a 5% error rate in latitude and longitude rounding, which could introduce small but meaningful inaccuracies in precise geospatial applications. They also struggled with date conversions, with a 10%–15% error rate in formatting timestamps correctly into ISO 8601. Even simple unit conversions, such as feet to meters, resulted in a 5% error rate, highlighting a consistent challenge in numerical data transformations.

To achieve 100% content accuracy, our findings indicate that additional validation and consistency checks must be incorporated into the processing framework. These include enforcing strict unit conversion rules, leveraging semantic matching techniques for natural language descriptions, and integrating knowledge graphs to improve structured data reconstruction when jargon and ambiguous terminology are present.

Conclusions and Future Work

This research addresses the challenges of interacting with complex military software systems by proposing a novel approach to conversational user interfaces. Our experiments demonstrate that while format similarity in structured data reconstruction is easily achievable with state-of-the-art LLMs and LLM frameworks, content accuracy remains a challenge, especially when processing natural language with jargon and domain-specific terminology. We found that KGs significantly improve accuracy in mapping ambiguous natural language descriptions to structured code values, particularly for fields like vessel type and vessel status. However, issues such as geospatial rounding errors, incorrect date formatting, and inconsistent unit conversions highlight the need for additional validation mechanisms to ensure high-fidelity structured data extraction.

As a next step, we aim to refine our approach by integrating verification techniques for numerical transformations, expanding the knowledge graph to cover more maritime-specific terminology, and fine-tuning LLMs with domain-specific training data. Further research will explore hybrid AI architectures that combine LLMs, KGs, and (potentially) deterministic validation mechanisms to achieve near-perfect content accuracy. These improvements will enhance the reliability of AI-driven structured data extraction, making it more applicable to real-world military simulators and other structured data domains.

References

Aghaei, R. K. (2025). *The potential of large language models in supply chain management: advancing decision-making, efficiency, and innovation*. ArXiv preprint arXiv:2501.15411.

Anurin, A. N. (2024). *Catastrophic cyber capabilities benchmark (3CB): Robustly evaluating LLM agent cyber offense capabilities*. ArXiv preprint arXiv:2410.09114.

Biswas, S. (2023). Prospective role of chat GPT in the military: According to chatGPT. *Quios*.

Caballero, W. N. (2024). *On large language models in national security applications*. ArXiv preprint arXiv:2407.03453.

Connolly, B. J. (2024). Battlefield information and tactics engine (BITE): A multimodal large language model approach for battlespace management. *SPIE Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*, *13051*.

Gao, Y. Y. (2023). *Retrieval-augmented generation for large language models: A survey*. ArXiv preprint arXiv:2312.10997.

Goecks, V. G. (2024). COA-GPT: Generative pre-trained transformers for accelerated course of action development in military operations. *International Conference on Military Communication and Information Systems (ICMCIS)*.

Hogan, A. E. (2021). Knowledge graphs. *ACM Computing Surveys (Csur) 54*(4), 1–37.

Javaid, S. F. (2024). Large language models for UAVs: Current state and pathways to the future. *IEEE Open Journal of Vehicular Technology*.

Kress, M. M. (2023). *AIS data: An overview of free sources*.

Logan, S. (2024). Tell me what you don’t know: Large language models and the pathologies of intelligence analysis. *Australian Journal of International Affairs*, 220–228.

Lukens, S. (2023). Evaluating the performance of chatgpt in the automation of maintenance recommendations for prognostics and health management. *Annual Conference of the PHM Society*, *15*(1), 1–18.

Makovec, B. R. (2024). *Preparing AI for compliance: Initial steps of a framework for teaching LLMs to reason about compliance*.

Mikhailov, D. I. (2023). *Optimizing national security strategies through LLM-driven artificial intelligence integration*. ArXiv preprint arXiv:2305.13927.

Nitzl, C. C. (2024). *The use of artificial intelligence in military intelligence: An experimental investigation of added value in the analysis process*. ArXiv. preprint arXiv:2412.03610.

NOAA. (2024). *AccessAIS*. https://coast.noaa.gov/digitalcoast/tools/ais.html

Olena, K. (2024). *Application of LLMs for a chatbot system in the logistics industry.*

Peng, C. X. (2023). Knowledge graphs: Opportunities and challenges. *Artificial Intelligence Review, 56*(11), 13071–13102.

Shafee, S. A. (2024). Evaluation of LLM-based chatbots for OSINT-based cyber threat awareness. *Expert Systems with Applications*.

The Use of Modeling Systems and Systems of Systems for Acquisition and Test in the DoD

Craig Arndt—Principal Research Faculty, Georgia Tech Research Institute. [Craig.Arndt@gtri.gatech.edu]

Sarah Scheithauer—Senior Research Engineering, Georgia Tech Research Institute. [Sarah.Scheithauer@gtri.gatech.edu]

Abstract

The Department of Defense (DoD) acquisition system is in the process of digital transformation. This effort is impacting all aspects, and areas of the acquisition system and also the different systems and process that interface with the acquisition system, including requirements development, test, operations, and threat analysis. Each of these processes have one or more key stakeholders. In each case one or more of these stakeholders generates models in order to facilitate the processes of acquisition (requirements, design, development, test, etc.). The proliferation of models is a good thing form the perspective of digital transformation and the training and cultural transformation of the workforce, however, in order to gain the full benefit from digital engineering and Model Based Systems Engineering these different models need to be coordinated and linked together in meaningful ways.

There are a number of not only different models in use in DoD acquisition and test, but a number of different classes of models which are used for different tasks and implemented using different technologies. Many of these models and modeling systems, were not originally designed to work with the other models that have been developed in other areas of the acquisition process.

The driving for a coordinated approach to developing not only the models effectively but also to prioritize the development of models that easily interface with each other come for the need for programs to be more efficient and the need to deliver capabilities to the war fighter faster. The coordination and integration of different models holds the promise to make significant improvements in these areas.

This paper addresses a number of the issues that arise from the development of a large number of disconnected models and systems. We identify specific areas for technical and for policy development and introduce and specific method for prioritizing work to grow the integration and coordination of these different models and systems.

Keywords: Model Based Systems Engineering, Digital Transformation, Program Management, Lessons Learned

Introduction

The objective of the Defense Acquisition System (DAS) is to support the National Defense Strategy, through the development of a more lethal force based on U.S. technological innovation and a culture of performance that yields a decisive and sustained U.S. military advantage. The acquisition system will be designed to acquire products and services that satisfy user needs with measurable and timely improvements to mission capability, material readiness, and operational support, at a fair and reasonable price. Within the DAS the development and fielding of defense systems is a complex process guided by and wide range of rules and processes. One of the most important processes in the system engineering process. In recent years new technologies have become available to improves the systems engineering process, specifically Model Based Systems Engineering (MBSE).

Digital Transformation

In June 2018 the Department of Defense (DoD) established its expectations for digital transformation in The DoD Digital Engineering Strategy. The strategy outlines five goals aimed at establishing a digital engineering environment for more rapid and effective development and fielding of weapon systems. The goals include using models to inform decision making, establishing an infrastructure to enable the digital engineering methods, and transforming the workforce to adopt digital engineering methods over the acquisition life cycle. Figure 1 was developed by the DoD to help communicate the different elements of the transformation effort. The development and use of standardized models is critical to the success of the transformation and the resulting advantages of digital engineering to the operations of all aspects of the department.

The DoD followed up this strategy with the release of formal guidance via DoD Instruction 5000.97 which ensures that the Director of Operational Test and Evaluation (DOT&E) will utilize digital engineering methods to achieve their test objectives for operational assessment and Live Fire Testing. Also in 2023, DOT&E released their DOT&E Strategy Implementation Plan (I Plan) which includes objectives and key actions to develop digital, or model based Test and Evaluation Master Plans (TEMP) and Integrated Decision Support Keys (IDSK). As recently as December 2024, the department has released an update to DoD Instruction 5000.98 and five DoD manuals further refining the description and use of digital methods for the entire DoD test community (DoDM 5000.96, DoDM 5000.99, DoDM 5000.100, DoDM 5000.101, and DoDM 5000.102).[1]

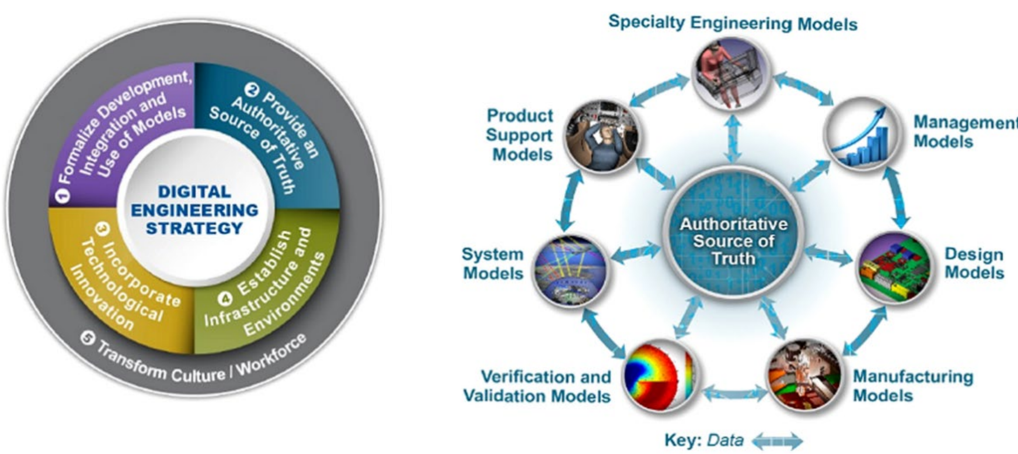


Figure 1. DoD Digital Transformation

Proliferation of Different Systems and Software

The are many different software systems that support digital engineering in use throughout the DoD and the vendor base that develops system for the DoD. When talking about Digital Engineering (DE) and MBSE in the context of the DoD, it is important to remember that both DE and MBSE have been commonly used in other industries for years before the adoption of these technologies by the DoD. As a result, there is a wide range of systems available in the market place for different organizations within the larger DoD ecosphere to get digital engineering infrastructure.

[1] https://www.dote.osd.mil/LinkClick.aspx?fileticket=Dt45nHpTB6A%3d&portalid=97

Background

As program offices and other organizations start to implement MBSE and other digital processes on programs they are discovering a wide range of implementation issues and complications.

Use of MBSE in the Defense Industry

The use of MBSE processes and tools is not new in the engineering industry nor the DoD vendor base. This has great advantages, in that the DoD is not starting from scratch with the implementation of a new technology. However, the fact that MBSE did not start in the DoD also presents significant challenges. These challenges include the fact that the SysML and many Product Lifecycle Management (PLM) systems are designed for commercial manufacturing and production systems, and often don't take into account the intricacies of the DoD acquisition process, such as the types of information and data exchange associated with DoD contracts.

PLM Tools

In industry, PLM is the process of managing the entire life cycle of a product from its inception through the engineering, design, and manufacture, as well as the service and disposal of manufactured products. PLM integrates people, data, processes, and business systems and provides a product information backbone for companies and their extended enterprises. In the case of the government the PLMs are chosen in order to facilitate the management of both government and vendor models needed to manage the development of new systems.

A wide range of software tools have been developed to support a product's life cycle. These include for example the Siemens Teamcenter. Teamcenter is a modern PLM system that connects people and processes, across functional silos, with a digital thread for innovation. The Teamcenter platform is primarily designed to support the design and development of products that subsequently get manufactured. The proliferation of PLMs and the lack of coordination between the different proprietary systems has proven to be a point of difficulty when integration between vendors and different government systems.

Development of MBSE Standards in the DoD

In the long-term however there is a great deal of need for both technical standards and processes. There are a number of different standardizing efforts currently on-going throughout different parts of the DoD and the extended defense industry. However, many of the standardization efforts have not been coordinated industry wide. Other outstanding issues with standards are: 1. The lack of a standard ontology, that is accepted across the industry, and 2. A lack of specific use cases that are used to verify the usability of the standards.

SysML Tools and Versions

In addition to issues with different standards and the lack of standards within the different parts of the industry there is the issue of the fact that the there are several different versions of the primary modeling language for MBSE, SysML. These different versions are not all compatible with each other and different stakeholders and developers use different versions of SysML. This creates considerable expense in conversion and incompatibilities that need to be resolved between organizations that need to transfer or deliver models.

Examples of Other Development

During the past few years there have been significant development programs created to standardize many key parts of the modeling process as it directly applies to DoD acquisition. Some of the more important activities are referenced below.

Arizona Ontology

In an attempt to address the issue of significant differences in both terminology and definitions of terms, but more importantly a differences in architectures, the University of Arizona developed a set of ontologies. These ontologies fill an important need for consistent definitions and architectures. Figure 2 shows how the ontology can be integrated with other acquisition related models. In this case the ontology is integrated with a Model Based TEMP.

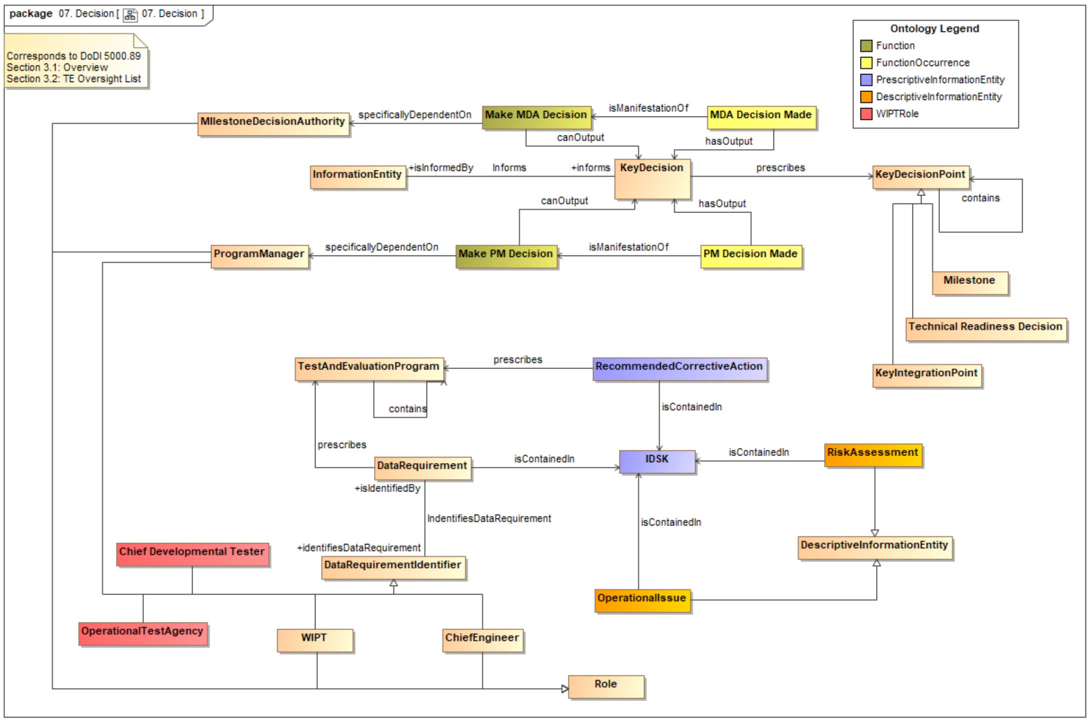


Figure 2. The Arizona Ontology as Implemented in SysML

Johns Hopkins University Meta Model

As a part of a project to improve the ability of different models to pass data and otherwise communicate Johns Hopkins University (JHU) teamed with the DoD to develop a Meta Model of MBSE process within the DoD acquisition process. Figure 3 shows a part of the Meta Model that will be used by the DoD to develop interface to allow future integration of models.

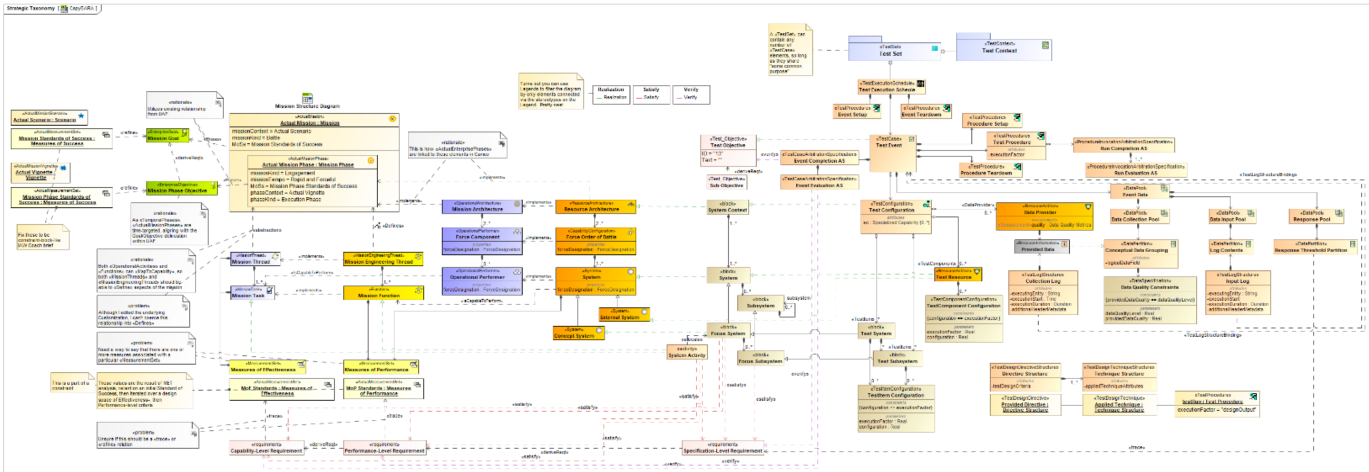


Figure 3. JHU Meta Model

Integrated Decision Support Key Architecture

The Integrated Decision Support Key (IDSK) was developed in order link acquisition decision making to the sources of data needed to make these decisions. To make the IDSK compatible with MBSE, DOT&E teamed with Georgia Tech to development the Model Based IDSK Reference Architecture. In Anyanhun and Arndt (2024) an MB-IDSK reference architecture (MB-IDSK RA) was proposed and developed to support digital transformation efforts of DOT&E. The motivation behind defining a MB-IDSK RA was based on the premise that an architecture should reflect the organization of the owning enterprise (CAS, 2022). Specifically, the MB-IDSK RA represents an essential tool to facilitate communication and alignment efforts of current and future IDSK architectures. Figure 4 depicts the IDSK architecture strategy as adapted from the DoD Comprehensive Architecture Strategy.

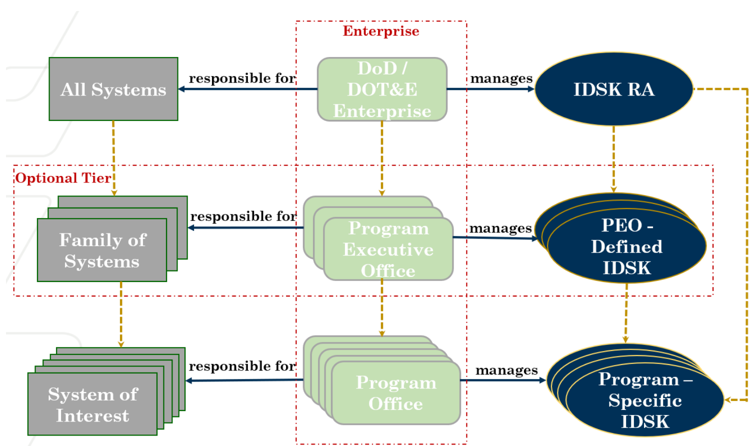


Figure 4. IDSK RA Architecture Strategy Adapted from Figure 1 of the DoD CAS (CAS, 2022)

The MB-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.

Model-Based TEMP Reference Architecture

The Test and Evaluation Master Plan or TEMP is a foundational document or artifact. The Model-Based Test and Evaluation Master Plan Reference Architecture (MB-TEMP RA) Model was developed using a domain-based approach. The MB-TEMP RA in an example of how more than one different model can be tied together for a common purpose architecturally. These models linked together within the Model Based TEMP include a Model-Based IDSK Reference Architectures, mission models, test range and facility models, test models, requirements models, system models, and the Test and Evaluation (T&E) Reference Metadata Model (2011).[2] An example of the architectural views of the MB-TEMP RA is portrayed in Figure 5.

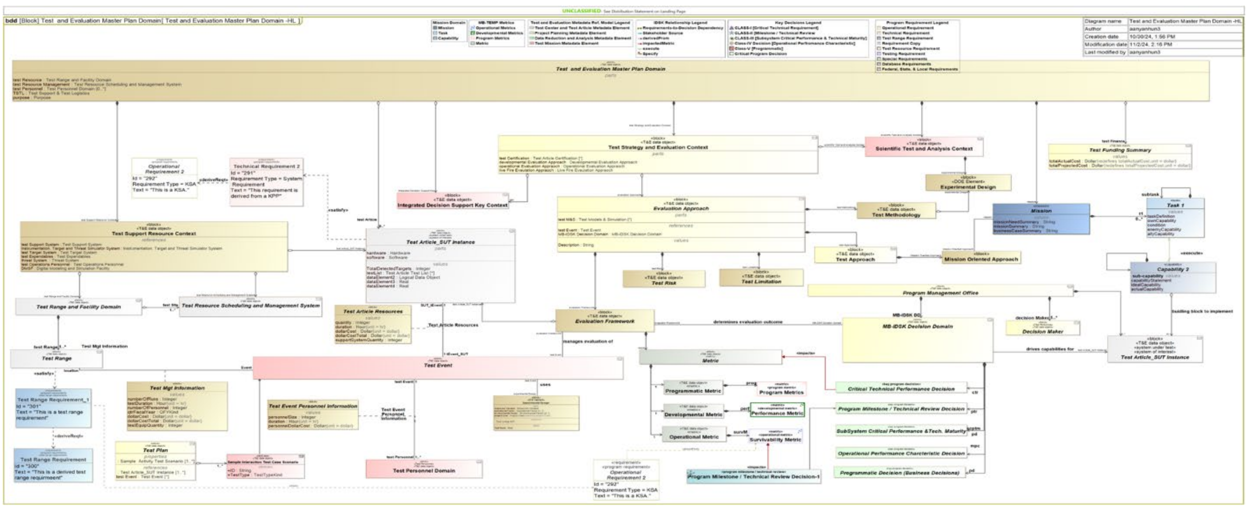


Figure 5. TEMP Domain View of the MB-TEMP RA

Figure 5 shows the TEMP Domain view of the MB-TEMP RA provides crucial insights into the top-level composition of the TEMP domain. The RA view links together elements defined within the TEMP model and elements already defined in digital models that exist within a program's digital engineering ecosystem. These digital models include a *program office model, requirements model, system model, SUT model, and test range models*.

Need

The DoD digital transformation gave a significant amount of guidance on performing the digital transformation. Figure 1 shows the top level of guidance in the DoD level transformation. However, this guidance does not provide guidance on the functional elements of the operations of digital acquisition.

In order to better understand what is needed for an end-to-end life cycle digital acquisition program we need to define a digital thread that looks at the different elements of both the acquisition program and the models that are involved on executing that program. This digital thread can be defined across the life cycle and also across the different models that will create a link between the models to allow for visibility of the data created across the different models to inform the other models and critical decisions that need to be made with regards to the system development. As can be seen in Figure 6, there are a wide range of different models created in the development of a new system. When the integrated digital thread is developed

[2] https://apps.dtic.mil/sti/tr/pdf/ADA640532.pdf

linking all of the different models together the data needed to make critical decisions more available to decision makers.

In commercial industry the process is much easier, due to one organization controlling the development of most if not all of the different models. This however becomes much more difficult in the case of the DoD, where the different models are developed by different organizations within the government and outside of the government, and no one has control over all aspects of the total set of models that need to be integrated together.

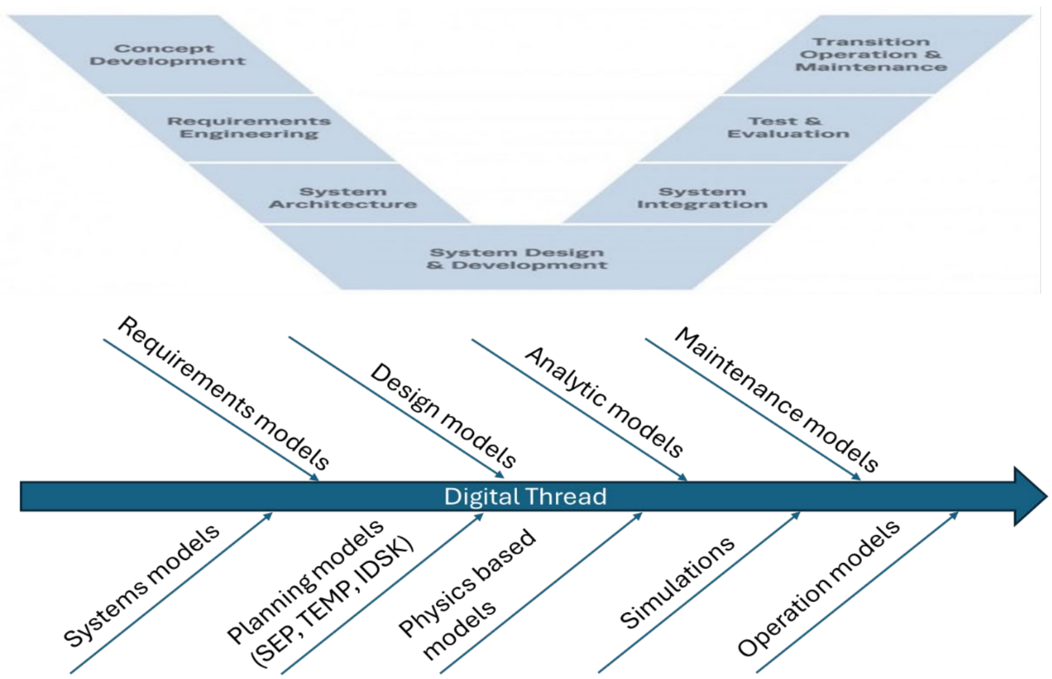


Figure 6. Digital Thread Across the Life Cycle

Progress in Many Places

The DoD community has been development many different digital tools and solutions for the acquisition process. The community involved in these developments has included, vendors, tool makers, government organizations, FFRDC's, and universities.

Communities of Interest

In the development of DE tools and MBSE tools several important organizations have been instrumental in the development of underlying constructs, and principals. INCOSE and the Object Management Group, or OMG, jointly chartered the OMG Systems Engineering Domain Special Interest Group or SE DSIG to create the Unified Modeling Language (UML) for Systems Engineering Request for Proposal which was completed in 2003. The development of the SysML modeling language is documented in Figure 7.[3]

[3] SysML History, https://www.incose.org/

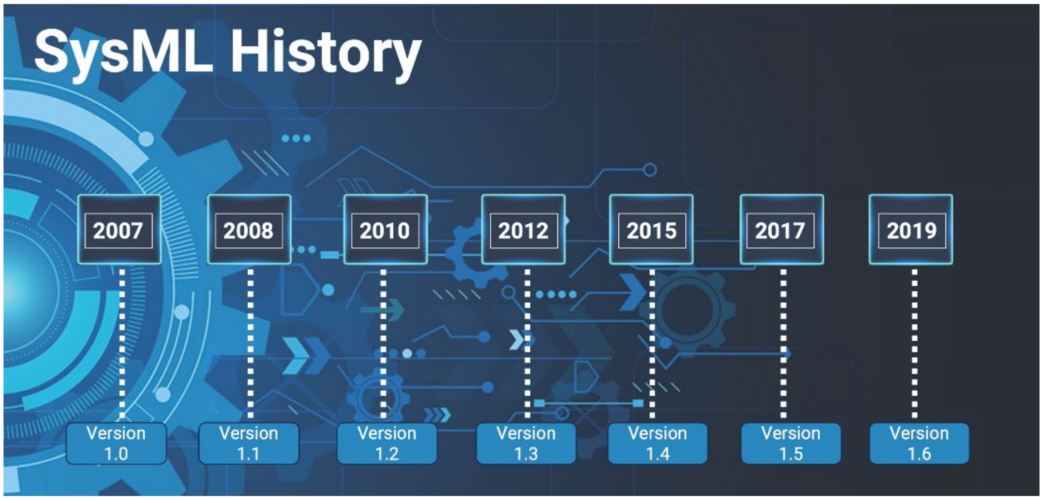


Figure 7. History of the SysML Language

DoD Policy Shops

In addition to industry groups, the leadership of the DoD have been instrumental in the development of requirements for the digital engineering tools that we use today. Several specific efforts have been forwarded by specific part of the DoD, and the Office of the Secretary of Defense (OSD).

DOT&E

The leadership of the DOT&E has done a great deal of work to define important acquisition artifacts in digital formats including the MB-IDSK and the MB-TEMP. Other groups, including the Office of the Undersecretary of Defense for Research and Development, Research and Engineering (OUSD[R&E]) have sponsored work in defining meta models of the acquisition process and structured data exchange metrics to accommodate the larger multi model environment.

Specific Program Offices

In addition, a number of forward-thinking program offices have chosen to be pathfinders in the development and implementation of digital engineering within their programs. These pathfinder programs have developed a number of key guidelines for the development of practical aspects of the MBSE within the programs. These programs have also contributed significantly to lessons learned.

Governance Issues

Governance can many different forms. When we talk about the governance of models, we look at sources for the different aspects of the governance: 1. direction about the structure of the models, 2. the content of the models, and 3. the interfaces and use of the other models to support the primary system models. At the highest-level international standards bodies (International Standards Organization [ISO] and Internation Council on Systems Engineering [INCOSE]), maintain top level standards for the SysML language. In addition to that, there needs to be standards that are specific to the defense industry and DoD acquisition.

DoD acquisition programs get the majority of the governance for DoD Instructions, and other policy documents. Traditional additional governance has been provided by panning document like the Acquisition Strategy, the TEMP, and the modeling and simulation plan, for the program. At this point additional governance is needed to deal with the complexities of managing the models needed to run a complex defense acquisition. One of the biggest complications in developing good comprehensive governance is the fact that many of the supporting models that are needed to develop the full life cycle of the program are not controlled by the program office or even the same government agency. Some governance can be provided to the vendors, though the RFP, and subsequently thought the contract, but outside government organizations that provide data and models that are needed to support the program and span the life cycle, including, threat, requirements, test, and configuration models. It is clear that comprehensive governance needs to be provided all stakeholders in order to ensure that content and interfaces needed to manage programs can be generated effectively.

Experience on real programs has shown us that the development of governance early is critical, because it needs to be provided to all the key stockholders before they develop their models.

Ongoing Challenges

With all of the work and advancements in digital engineering and MBSE, there still a number of major challenges that need to be both better understood and overcome in order to fully realize the potential of DE and MBSE.

Integration Across the Life Cycle

There are many different classes of models, and different models, that will need to be integrated, in order to make an integrated system of models. There are many specific challenges to integrating these different models. The integration of the different models will require dedicated interfaces.

The power of MBSE is multiplied when data can be shared across time and across all of the different models that are developed by different organizations involved in the development, fielding and operations of the system. To achieve a greater level of integration of the different models across time and models, will require: 1. better documentation of data formats and structure in all of the relevant models, 2. better version control for supporting software systems and languages, and 3. some degree of coordination between the developers of the different models.

Return on Investment and Measures of Effectiveness

Currently there are a wide number of possible levels of implementation of MBSE within specific programs and different parts of the development life cycle. However, there is a limited amount of time available to conduct meaningful return on investment (ROI). In order develop these systems in a meaningful way, we will need to effectively develop metrics to evaluate progress.

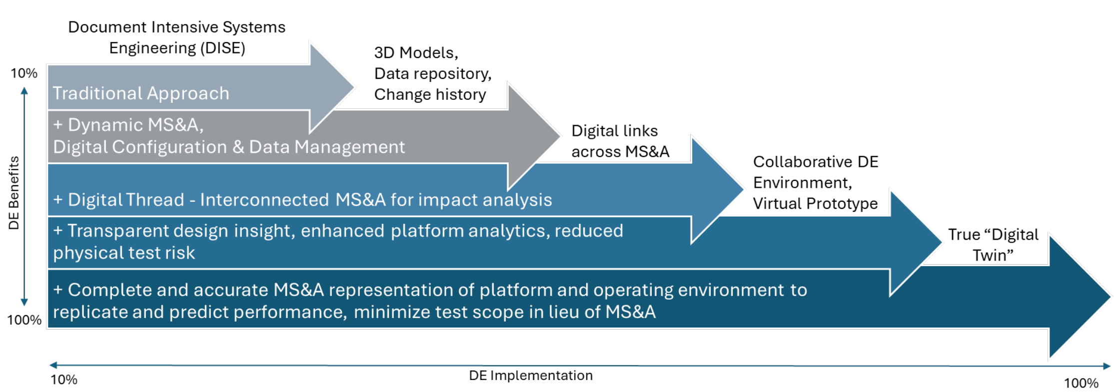


Figure 8. A Structured Approach to Different Levels of Implementation

However, there is a lack of tools and methods available in order to make good decisions on the level of implementation based on the return on investment for specific programs and meaningful metrics for the return in investment. There are any number of possible ways to implement digital engineering into any given program. Figure 8 shows the incremental approach that is being look at by many DoD programs. This allows programs to select the pathway to digital implementation that make the most sense given their current state and the resources they are willing to expended to make the conversion to digital engineering.

This is however only the first step in implementing a fully digital life cycle for any given system. The issue remains that many of the vendor systems do not match systems and processes adopted by the program offices. The adjacent and supporting processes and models that support and / or feed into the program offices models and process are not coordinated or compatible with the program offices' systems. This lack of coordination continues to create significant issues for the program offices.

Conclusions

To say that we need to better integrate the different models involved in the acquisition process, is true, but does not tell the complete story of what is needed to facilitate realizing the benefits of DE. The full realization of an integrated digital modeling environment will need to be achieved incrementally for a wide range of reasons that we have discussed here. The difficulties the DoD is having with digital transformation are both technical and also programmatic. Several key things that we have discovered about these issues are summarized below. One of the first things that needs to be done is to ensure that we are learning about incompatibilities and technology disconnects before it is far too late in the development programs. After a contract is awarded and on a fixed timeline it is far too late to discover incompatibilities.

Ideally, the DoD's digital integration system will be transparent across different models and systems. There have been a number of different technical issues that have prevented this from happening to the degree that could be possible. Some of their technical courses have included differences in ontologies, PLMs and other non-modeling IT resources, and infrastructure.

Non-technical issues include programmatic / contracting issues. The introduction of DE has proven to be a challenge to all aspects of the acquisition process. Contracting for digital deliverables requires a number of significant differences in the contracting process, including new Digital Data item Descriptions (DiDs). The acquisition community and the Defense Acquisition University (DAU) and others have been collecting lessons learned from a wide range of sources. These efforts need to continue and be expanded, shared, and collated.

Recommendation / Path forward

There is a lot that we can, should, and need do to accelerate the digital transformation and implementation of MBSE. The DoD has chosen to adopt DE practices within the context of the DoD priority to accelerate the delivery of systems to the field. Below are several key next steps that can be done to accelerate the DoD's ROI in terms of accelerating the acquisition.

Working More and Better with Vendors

As we have seen in different parts of this paper the interactions between the program offices and vendors can be very complex, difficult, and time consuming. Considerable ROI can be achieved by creating transparent connections between the government acquisition organizations and the different vendors developing programs.

More Standards and Reference Architectures

In order to guide the development and integration of different digital engineering, and more partially specific standards and methods (similar to IEEE/ISO/IEC 15288-2015) that will allow program offices to quicky and easily implement DE and MBSE process. These processes need to be structured into very specific and definable pathways for specific tasks and digital threads that lead to improved acquisition outcomes.

More Pervasive Training of All Parts of the Acquisition Work Force

Every major plan for digital transformation that the DoD and other organizations has strongly recommended training for the DoD workforces in DE and MBSE. Most of these recommendations have focused on the engineering teams that will be working with the vendors. To date, the DoD's efforts in training have not been as effective or as wide spread as they would have liked. In addition, focusing on the engineering teams has not advanced the goals of developing a full digital program. We recommend that the DoD increase the DE and MBSE training and expanded it to program management and contracting areas.

Reduction in Redundant Activities

Over the past few years there has been significant research and development in the area of applications of both DE and MBSE conducted in order to advance these areas. However, at some point this research and development becomes redundant when efforts are not coordinated. We have gotten to that point. In addition, the different research and development efforts are in many cases incompatible. In order to maximize the effectiveness of different research and development the efforts will need to be coordinated and reconciled based on guidance for the government.

References

Anyanhum, A., & Arndt, C. (2024, September). *The use of AI and systems engineering processes in and for testing of AI-based systems*. SERC AI4SE/SE4AI Workshop.

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

Beers, S. M. (2023). *Integrated decision support key (IDSK) for capability evaluation-informed MDO decision-making*. ITEA MDO Workshop, July 2022.

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

Director, Operational Test and Evaluation. (2017). *Test and evaluation master plan (TEMP) guidebook version 3.1*.

DoD. (2020). *DoD instruction 5000.89: Test and evaluation*.

DOT&E. (2022). *Office of the Director, Operational Test and Evaluation strategy update—Strategic pillars*.

https://www.dote.osd.mil/Portals/97/pub/reports/FINAL%20DOTE%202022%20Strategy%20Update%2020220613.pdf?ver=KFakGPPKqYiEBEq3UqY9lA%3d%3d

Executive Services Directorate. (2020). *DoD instruction 5000.89*. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500089p.PDF

Gregory, J., & Salado, A. (2024). An ontology-based digital test and evaluation master plan (dTEMP) compliant with DoD policy. *Systems Engineering.*

Gregory, J., & Salado, A. (2024). dTEMP: From digitizing to modeling the test and evaluation master plan. *Naval Engineering Journal, 136*(1), 134–146.

Gregory, J.& Salado, A. (2024). Towards a systems engineering ontology stack. In *INCOSE International Symposium*.

IEEE SA. (n.d.). IEEE/ISO/IEC 15288-2015: ISO/IEC/IEEE international standard—Systems and software engineering—System life cycle process. https://standards.ieee.org/ieee/15288/5673/

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

Muller, G,. & Hole, E. (2007). *Reference architectures: Why, what and how* [White paper]. Embedded Systems Institute and Stevens Institute of Technology.

OpenCAESAR. (2024). *Ontological modeling language 2.0.0*. [https://www.opencaesar.io/oml/

Range Commanders Council. (2011). *Test and evaluation metadata reference model*.

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.

Critical Skills for Successful Leadership of Large Complex Projects

Tom McDermott—is the Chief Technology Officer of the Systems Engineering Research Center (SERC) and a faculty member in the School of Systems and Enterprises at Stevens Institute of Technology in Hoboken, NJ. He previously held roles as Faculty and Director of Research at Georgia Tech Research Institute and Director and Integrated Product Team Manager at Lockheed Martin. McDermott teaches system architecture, systems and critical thinking, and engineering leadership. He provides executive level consulting as a systems engineering and organizational strategy expert. He is a fellow of the International Council on Systems Engineering (INCOSE). [tmcdermo@stevens.edu]

Nicole Hutchison—is a Research Associate Professor at Viginia Tech National Security Institute. Prior to that, Hutchison was a senior research scientist at the Systems Engineering Research Center. There, she collected the data and conducted analyses into skills, career paths, and personal enabling characteristics of systems engineers which comprised the Helix project. Hutchison is currently the Editor in Chief of the Systems Engineering Body of Knowledge. Prior to the SERC, she supported many branches of the federal government at analytic services (ANSER) in Virginia. She is a Certified Systems Engineering Professional and has more than 30 publications on systems engineering. [nlong@vt.edu]

Abstract

This paper explores the critical leadership competencies required for the successful management of megaprojects—large-scale, complex ventures exceeding $1 billion in cost and spanning multiple years. Given that only 8.5% of megaprojects are completed on time, within budget, and with intended benefits, effective leadership is vital. Successful megaproject leaders combine technical expertise with systems thinking, transformational leadership, and the ability to navigate complexity and uncertainty. They prioritize stakeholder management, communication, and people leadership, while maintaining a strong foundation in project management. Drawing from existing frameworks, including the HELIX systems engineering effectiveness model, the article proposes a new framework for developing the unique skills needed for megaproject leadership. This framework emphasizes the integration of technical, interpersonal, and enterprise leadership skills, which are crucial for managing the challenges inherent in these large, multi-stakeholder projects. The paper advocates for a shift in how leaders are selected and developed, focusing on their adaptability, strategic thinking, and ability to lead in uncertain environments. The goal is to improve the success rate of megaprojects through more effective leadership development and selection strategies.

Introduction

According to the Oxford Handbook of Megaproject Management, megaprojects are “large-scale, complex ventures that typically cost $1 billion or more, take many years to develop and build, involve multiple public and private stakeholders, are transformational, and impact millions of people” (Flyvberg, 2017). These projects are often mega-systems and are marked by operational uncertainty, behavioral complexity, pluralistic decision-making, and external environmental volatility (Stevens, 2011). Megaprojects, as with many large, complex systems, have a lower-than-desired success rate. Flyvbjerg, a leading scholar in megaproject management, suggests that only 8.5% of megaprojects are completed on time and within budget while delivering the intended benefits. This means that more than 90% of megaprojects fail in at least one key aspect: cost, schedule, or intended outcomes (Flyvberg, 2014). As megaprojects cost billions of dollars and often involve the labor and cooperation of tens of thousands of people, improving the success rate of these projects is critical. Megaprojects are often distinguished from other large projects in that they create “temporary enterprises” due to scale and uniqueness, implying the need for leaders who can quickly build and scale new enterprises from sets of existing enterprise structures in a complex supply network. Merrow, an authority in megaproject leadership, highlights that “the fate of difficult projects seems to hinge more on the project leaders than the results of simpler projects” (Merrow, 2011). In a study, co-authored with Nandurkdikar, Merrow examined 100 megaprojects and their leaders to determine the factors that differentiate between successful and less-successful leaders of large-scale, transboundary, complex projects. They found successful megaproject leaders spend more time on stakeholder management, communications, people management, and managing contractors and vendors; and spend less time on project processes, project controls, and engineering and technical tasks. However, they find project management is still important as a mastered set of skills. They also found that most successful megaproject leaders have engineering backgrounds (Merrow & Nandurdikar, 2018). This work found that the necessary intersection of engineering, project management, and megaproject leadership were most closely associated with systems engineering proficiencies.

A complex project leadership competency model developed by the International Centre for Complex Project Management (ICCPM) contends that the development of leadership skills follows a journey from Situational Leadership (leadership style adapts to the skills and needs of the team) to Contextual Leadership (leadership style adapts to the changing context of the business and its environment) to transformational leadership (leadership style motivates others to achieve shared vision and personal growth; ICCPM, 2021).This implies that successful megaproject leaders have evolved from engineering and technical program management roles to enterprise level roles and they have successfully combined technical leadership, enterprise leadership, and transformational leadership.

Our research focused on megaproject leaders across government and industry led projects. We found a strong correlation between past research on the proficiencies of effective systems engineers in the HELIX project (Hutchison et al., 2020), and as applied to technical leadership development (Pennotti et al., 2015) in earlier career stages. Both frameworks were developed by the SERC. This research used literature review and discussion with megaproject leaders to understand how successful megaproject leaders might evolve and be developed. We provide a new framework that details a set of unique competencies that are required for successful megaproject leadership, how these competencies can be distributed amongst a team, and how organizations seeking to initiate mega projects should think about the selection of the leadership team. Many of these competencies have foundations in systems engineering.

Existing Frameworks with Relevant Skills and Competencies

There are several existing frameworks that have relevance for the creation of megaproject leadership. If megaprojects are also mega-systems, systems engineering frameworks have relevancy, as do frameworks that deal with complexity and project and technical leadership.

Helix—HelixSE (systems engineering) resulted from research from the SERC that examines what makes systems engineers effective. As megaprojects are also mega systems, systems engineering skills are important inputs to megaproject leaders (Hutchison et al., 2020). The Helix framework is shown in Figure 1.

HelixEMP (employability skills) resulted from additional studies by SERC researchers Hutchison and McDermott, taking the foundational framework of HelixSE and expanding it to cover employability skills more generally. There are six sets of characteristics that emerged: foundational learning, domain knowledge and experience, systems knowledge and experience, self-leadership and learning, team leadership and collaboration, and complex problem-solving (McDermott et al., 2021). These provide the underpinning for the development of transformational leaders early in their careers and will be shown to form the nexus of our megaproject leadership competencies framework, and the core of the early-stage learning for megaproject leadership.

Figure 1. HELIX Proficiencies of Effective Systems Engineers (helix-se.org)

Technical Leadership—Researchers from the SERC worked with leaders in the Defense Acquisition University (DAU) to develop a framework for explaining the skills required for effective technical leadership (Pennotti et al., 2015). This utilizes three lenses, which build in complexity: systems, business, and enterprise, as shown in Figure 2. The framework focuses on personal leadership strengths and competencies with a systems lens; team and group leadership skills necessary to execute a business; and the enterprise leadership skills necessary to enact change and move forward. As megaprojects are evolutionary, a key aspect of this framework is managing emergence and evolution of the system, in the context of business and enterprise change strategies.

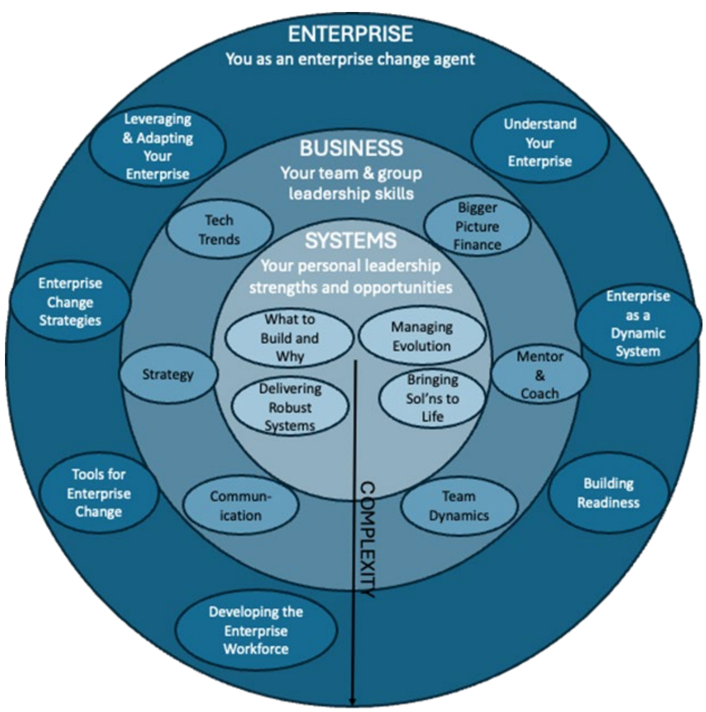


Figure 2. Overview of the Technical Leadership Skills Outlined in Pennotti et al. (2015)

Bass Transformational Leadership Theory—Bass defined two transactional and four transformational characteristics that contribute to high performance in leadership (Aarons, 2006). While transformational leadership motivates and inspires followers, transactional leadership is more focused on “exchanges” between leader and follower in terms of work tasks, penalties, and rewards (Van Wart, 2015). Van Wart further tested a framework in his case study of transformational leadership. This framework is important to megaproject leadership because both transactional (project management) and transformational (influence and followership) leadership skills are necessary. The two transactional characteristics are more related to the project plan and include: management by exception (deal with low-performing employees, avoid technical mistakes and blunders, deal with performance lapses, stabilize organization if needed) and contingent reward (pay, work-life balance). The four transformational characteristics are more motivational and include: individualized consideration (coaching, delegation, training opportunities), idealized influence (model exemplary behavior, avoid personal scandal, use charismatic communication), inspirational commitment (gain commitment to the work and the profession, teamwork), and intellectual stimulation (ensure the need for change, provide a plan, build internal support, ensure top management support, ensure external support, provide resources, institutionalize changes, pursue comprehensive change; Aarons, 2006). Figure 3 depicts how these characteristics come together in a project setting.

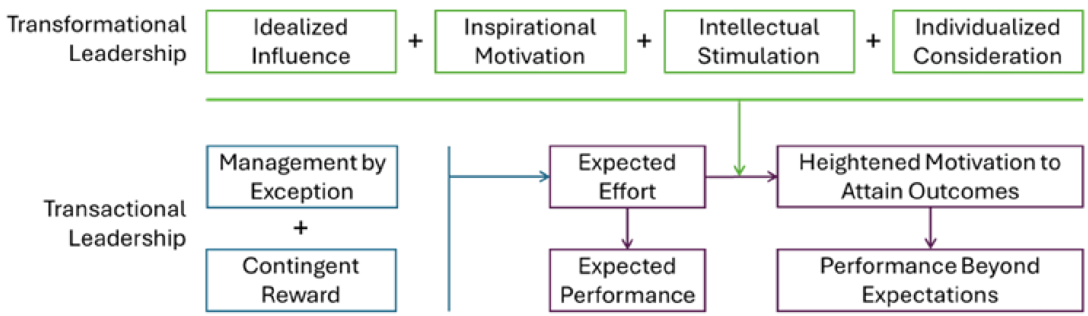


Figure 3. Bass Transformational Leadership Model (Aarons, 2006)

Complex Project Leadership Competency Standards—The International Centre for Complex Project Management (ICCPM) generated these standards via a collaborative effort of 169 individuals from 24 nations to identify the critical skills of complex project leadership. The framework addresses abilities to navigate complexities, foster innovative thinking, manage risks, and inspire high-performing teams. The framework integrates across three leadership theories: situational leadership, contingency theory of leadership, and transformational leadership. We found the career journey from situational leadership to transformational leadership is at the heart of megaproject leadership development—an integration of the HELIX and ICCPM frameworks. The ICCPM framework is divided into five main competency areas: drive systemic thinking and action (how leaders address complexity, uncertainty, and emergence), focus strategically on delivering project outcomes (the elements that lead planning and execution of megaprojects), engage collaboratively with stakeholders (strategic conversations with critical stakeholders), exercise contextual leadership awareness (self-aware), and apply system governance and delivery assurance (evolve the project to assure success; ICCPM, 2021).

Literature on Qualities and Skillsets of Megaproject Leaders

Any review of literature on megaproject leaders must start with Merrow and Nandurdikar’s book, *Leading Complex Projects* (Merrow & Nandurdikar, 2018). This book provides deep insights into managing large-scale, complex projects, although focused solely on the oil and gas industry. The book discusses the unique challenges of such projects, emphasizing the need for strong leadership, effective communication, and rigorous planning. Key themes include managing uncertainty, stakeholder engagement, decision-making under pressure, and balancing short-term objectives with long-term goals. The authors offer practical strategies based on research findings to help project leaders improve performance, mitigate risks, and navigate the complexities inherent in large, high-stakes projects. Historically, excellent program managers have been selected as megaproject leaders. But success in lower-complexity projects is not indicative of successes in more complex projects. “It is not even clear that a long career progression in smaller, less-complex projects is of any substantial value to the complex project leader at all.” Skillsets for PMs overlap with megaproject leadership—but there are very different emphases. Most effective megaproject leaders have had:

- Broad and varied jobs. Most successful megaproject leaders have depth in the domain of the project but have worked more broadly and become generalists.
- Experience making timely decisions under conditions of uncertainty.
- Opportunities to watch others lead complex projects.
- Experience sorting out difficult interpersonal situations.
- Backgrounds in engineering (most megaprojects are engineered systems).

Critical skills for megaproject leaders as cited by Merrow and Nandurdikar (2018) include:

- Highly open to new experience, self-disciplined, engaging, stable, and test high in emotional intelligence (the five-factor model)
- Project management remains important but in the context of cooperation and not in transactional methods (as it is usually employed)
- Preference for spending time on people management, alignment, and communications over on work processes
- Highly aware of their own abilities for learning

This is the most comprehensive resource in the literature. In addition, more than 70 articles on megaproject leadership skills were reviewed and used to produce the framework in the next section.

Framework for Megaproject Uncertainties

Overall, several key themes appeared consistently across megaprojects literature. The foremost of these were the ability to manage complexity and project uncertainties (as opposed to risks). The research built an uncertainty framework to characterize types of uncertainties in megaprojects and strategies to deal with them, shown in Figure 4 (McDermott et al., 2024). This led to a playbook of project strategies across eight areas of uncertainty.

- In the Strategic Context, megaprojects are characterized by more **complexity** and **uncertainty** in external environment, context, and mission; and more complexity and uncertainty in their internal spans of control.
- In the System Context, megaprojects tend to be more **transformative** in the system outcomes; and more transformative in their processes associated with predictability of core concepts. Megaprojects need to plan and execute more **flexible** decision-making processes.
- In the Implementation Context, megaprojects tend to incorporate more novelty and innovation and have less knowledge of end design; and less knowledge of cross-domain relationships that create complexity. Projects should invest in **flexibility** to manage **risk**

and **uncertainty**, particularly modularity so that “unknowns” can be separated from “knowns.”

- Finally in the Stakeholder Context, megaprojects tend to have less alignment of and trust between **stakeholders**; and need more strength in stakeholder relationships. Projects must focus more on maintaining and sharing project knowledge.

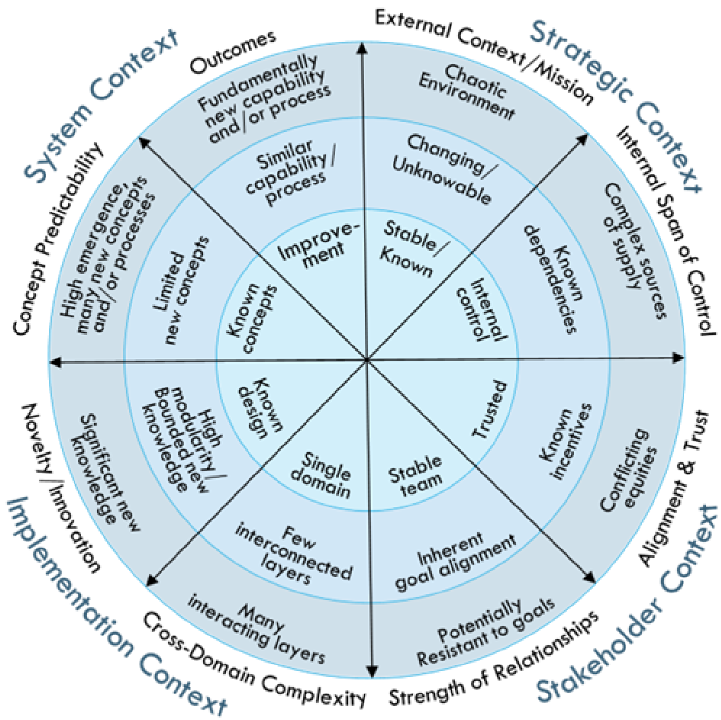


Figure 4. The Megaproject Uncertainty Framework

In particular, there were five major leadership and management capabilities that stood out as necessary for the successful completion of megaprojects. These are the abilities to **1) Manage Uncertainty, 2) Manage Complexity, 3) Lead Transformation, 4) Manage Stakeholders, 5) Create Flexibility, and 6) Manage Risk**.

Framework for Megaproject Leadership Skills

The critical skills and abilities (competencies) for megaproject leaders were identified based on the literature but also with these six abilities in mind. Each of the competencies outlined in the framework are necessary to successfully perform each of these six activities. Stated differently, it takes a number of individual competencies to deliver these capabilities effectively. We identified the following major competency areas required for megaproject leaders: **domain/discipline foundations**, **personal characteristics**, **mindset**, **thinking skills**, **interpersonal skills**, and **enterprise technical leadership**. Each of these six areas is defined and explained in the following subsections. Note that “mindset” and “thinking skills” are related. In general, mindset is about the attitudes and beliefs that influence how someone approaches situations while thinking skills are the cognitive tools and processes used to analyze and solve problems. These areas are related, and these skills go together, with mindset often shaping how effectively thinking skills are applied.

Domain/Discipline Foundations

Historically, technical leaders have been more effective when they have had deep experience in the primary domain and at least some experience with the most critical disciplines for the system. This is true at smaller scales of complex projects as well as at the megaproject level (Pennotti et al., 2015). Merrow and Nandurdikar's analysis of 100 megaproject leaders highlighted that most of the successful leaders had a technical background, often in engineering (Merrow & Nandurdikar, 2018). This does not mean that megaproject leaders should do engineering work, but that they should have enough technical depth to be able to make major technical decisions, have effective dialogs on technical/engineering concerns, and know when to consult experts in a given discipline. One important note in Merrow and Nandurdikar's work is that leaders who have experience in more than one domain—but include experience in the primary domain for a megaproject—are more likely to be successful. The specific domain(s) and discipline(s) needed are dependent upon the individual megaproject. In the example of successful megaproject leader Dzulkarniain Azaman, Merrow and Nandurdikar highlight:

> "This early experience . . . is similar to what we have seen in the early careers of other leaders: A solid grounding in the fundamentals of their domain with exposure to the world of project management is followed by managing smaller-scale mini-projects and, more importantly, seeing projects from end-to-end, which provides the learning of seeing one's design, engineering, and project management decisions play out in the field." (Merrow & Nandurdikar, 2018)

Thus, the successful megaproject leader moves from a single discipline with deep technical background to mastery of multiple disciplines, while retaining technical depth in one or more disciplines. They also develop deep project management and systems engineering skills. In the process they develop skills across multiple project-related roles including management of large complex teams and supply chains. This in itself is not enough, as they also have to move from transactional to transformational leadership styles. The remainder of the framework outlines the development of skills necessary to make this transition.

Personal Characteristics

Though personal characteristics are not always included in competency frameworks, in the case of megaproject leadership, these characteristics emerged so strongly that it is critical to include them. In as much as these characteristics are more intrinsic/difficult to change, they become critically important for the selection of the right individuals for megaproject leadership. There are six critical characteristics for megaproject leaders: Self-Aware, Self-Motivated, Humble, Open-Minded, and Courageous.

Self-Aware: The ability to understand and recognize one's own thoughts, emotions, and behaviors, and how they affect oneself and others. Self-aware leaders are more likely to critically assess their own judgements and decisions. This can help prevent unrealistic planning and forecasting that can lead to project failure (Flyvberg, 2017).

Self-Motivated: The internal state that helps us initiate, continue, or terminate a behavior. Self-motivated leaders are more likely to exhibit a transformational leadership style, characterized by high levels of enthusiasm, passion, and a strong commitment to project goals (Eweje et al., 2012).

Trustworthy: The quality of being deserving of trust or confidence, often demonstrated through dependability or reliability. Followers make a leader and for that followers need trust in their leaders (Merrow & Nandurdikar, 2018).

Humble: A modest view of one's own importance; a lack of pride and arrogance. In the context of megaproject leadership, a humble leader is one who sees their role as serving others.

The ability to search actively for evidence against one's beliefs, plans, or goals, and weigh such evidence fairly (University of Pennsylvania [UPenn], 2024).

Open-Minded: Able to consider something without prejudice or application of pre-conceived notions. Leaders who are open to different ideas, flexible in their thinking, and willing to revise their plans based on new information tend to achieve better project outcomes (Merrow, 2011).

Courageous: The ability to do something difficult or risky, even in the face of fear or danger. It essential for leaders to confront the inherent risks, uncertainties, and complexities of megaprojects. Courageous leaders are needed to make tough decisions, address underperformance, and challenge unrealistic expectations or assumptions, especially when facing pressure from stakeholders or political forces (Doherty, 2008; Flyvberg, 2017).

Mindset

Mindset is about the attitudes and beliefs that influence how someone approaches situations. These are closely related to the thinking skills that are discussed in the following section. The critical characteristics of the mindset of successful megaproject leaders are Comfort with Uncertainty (the ability to function in an environment of unpredictability and lack of sufficient information); Paradox Mindset (the ability to hold and balance seemingly opposed views and being able to easily move from one perspective to another; Hutchison et al., 2020); Strategic Thinking (essential for leaders of megaprojects due to the scale, complexity, and long-term implications of these endeavors, this involves the ability to envision the future, anticipate challenges, align resources, and make decisions that drive the project toward its objectives despite uncertainties); and Vision/Goal Setting (vision being broad idea of what the megaprojects is trying to achieve in the future, while goals are specific aims or steps that will help reach that vision. The ability to create the vision and then identify key steps to attaining it is critical for megaproject leadership.)

Comfort with Uncertainty: The ability to function in an environment of unpredictability and lack of sufficient information. Leaders must embrace uncertainty and ambiguity to navigate large, complex projects effectively, using adaptive leadership styles (Doherty, 2008).

Paradox Mindset: The ability to hold and balance seemingly opposed views and being able to easily move from one perspective to another. Strategic vision paired with the ability to explore details, and judgement to know when more detailed exploration is appropriate and necessary (Merrow & Nandurdikar, 2018; Pennotti et al., 2015; Royse, 2021).

Strategic Thinking: Considered essential for leaders of megaprojects due to the scale, complexity, and long-term implications of these endeavors. It involves the ability to envision the future, anticipate challenges, align resources, and make decisions that drive the project toward its objectives despite uncertainties. Strategic thinking is required for understanding the full scope of risks, setting realistic goals, and ensuring alignment between project objectives and broader organizational or governmental strategies (Doherty, 2008; Hutchison et al., 2020; Merrow, 2011; Merrow & Nandurdikar, 2018).

Vision/Goal Setting: A vision is a broad idea of what the megaproject is trying to achieve in the future. Goals are specific aims or steps that will help reach that vision. A clear vision is crucial for guiding decision-making and maintaining stakeholder alignment in complex megaprojects (Flyvberg, 2017; Merrow, 2011).

Thinking Skills

Thinking skills are the cognitive tools and processes used to analyze and solve problems. These are related to mindset (previous section), with mindset often shaping how effectively thinking skills are applied. The critical thinking skills of successful megaproject leaders are Political Savvy, Recognizing Patterns, Solving Complex Problems, and Anticipating Future Situations.

Political Savvy: The ability to exhibit confidence and professional diplomacy, while effectively relating to people at all levels internally and externally (NIH). When teams believe the project is caught up in politics either from internal or external stakeholders, they can easily become distracted and discouraged. If the politics continually interfere, the team blames the project director and their confidence is undermined (Merrow & Nandurdikar, 2018).

Recognize Patterns: Abstraction is the quality of dealing with ideas rather than specific events. This is part of the ability to recognize patterns across different disciplines, domains, or contexts. Recognizing patterns in past projects is critical to anticipate potential pitfalls and optimize decision-making (Flyvberg, 2017).

Solve Complex Problems: Finding solutions to difficult or complex issues. Megaprojects are complex, involve a wide variety of stakeholders, and require the integration of multiple disciplines. By their nature, they will generate many problems which need rapid solutions. In complex problem solving, knowledge exchange and transfer within and across social networks is critical and knowledge visualization is a primary tool for developing shared understanding as a foundation for generating new knowledge.

Innovate: Coming up with new ways to do things; making changes to methods, processes, and approaches. Innovative approaches are critical for overcoming the inherent complexities and uncertainties of megaprojects (Flyvberg, 2017).

Anticipate Future Situations: The ability to predict what will happen or be needed in the future; foresight. This is closely related to the Mindset capability of Vision/Goal Setting. The ability to anticipate future situations is vital for navigating the complexities and uncertainties inherent in megaprojects. Leaders with foresight can better plan for long-term outcomes and adapt to evolving project environments (Flyvberg, 2017; Shenhar & Dvir, 2007).

Interpersonal Skills

By their nature, megaprojects involve a diverse set of stakeholders, both internally to the team and externally. Every reference in the literature review cited the importance of successfully managing stakeholders for the success of large, complex projects. These are closely tied to the competencies necessary for Enterprise Technical Leadership (following section). Success depends on proactive communication, trust building, and frequent engagement (Van Wart, 2015). A wide variety of interpersonal skills are required for megaproject leaders, though they can generally be grouped into the three main competencies: Communication, Relationship Building, and Negotiation/Persuasion.

Communication: Fundamentally, communication is the ability to convey and receive information effectively. There is a wide array of communication skills, deadline with how you express yourself, understand others, and adapt your communication style based on the audience. Successful megaproject leaders must master all aspects of effective communication. Communication is a focused task of successful complex project leaders (Merrow & Nandurdikar, 2018). Active listening is critical. Listening to understand versus listening to respond is a reflection of openness in communication style (Gil & Lundrigan, 2012). Ability to convey complex information clearly and succinctly (UPenn, 2024).

Relationship building: The process of creating and maintaining connections with people to create a sense of team or to achieve a purpose. Strong relationships enable better communication, risk management, and conflict resolution (Merrow, 2011). Communication must be at a higher, more visionary level in megaprojects (Merrow & Nandurdikar, 2018).

Negotiation/Persuasion: Megaprojects by nature have a diverse community of stakeholders. There will be conflicts between stakeholders and conflicts of priority within the team (e.g., trade-offs between budget, schedule, and performance). Managing stakeholders calls for negotiating and diplomatic skills (Merrow & Nandurdikar, 2018). Megaproject leaders need to be able to bring stakeholders together on critical decisions (Hutchison et al., 2020).

Enterprise Technical Leadership

By their nature, megaprojects are large, complex temporary enterprises. They require enterprise leadership skills in addition to the technical leadership skills required to oversee successful innovation. Enterprise Technical Leadership skills are: Promote Learning Culture, Foster Openness, Collaborate, Demonstrate Trust, Resolve Conflict, and Coach and Mentor.

Promote Learning Culture: Create an environment that encourages and prioritizes gaining new knowledge and skills. Learning is rewarded. The megaproject leader provides an example of continuous learning. Creating a learning culture in the context of megaprojects is critical to adapt to complexity, improve performance, and foster innovation. A continuous learning culture is necessary to address complexities and unforeseen challenges (Flyvberg, 2017).

Foster Openness: Encourage a culture of openness and communication and be open to new ideas and perspectives. This is closely linked with the individual leader's open-mindedness. Openness is a core competency for handling complex project dynamics, fostering transparency, and building trust. Openness, particularly in sharing information and acknowledging uncertainties, is key to building trust and achieving project success (Merrow, 2011).

Collaborate: The ability to work with others to produce or create something is a critical aspect of megaprojects. Building a collaborative team is critical to megaproject success. Collaboration across stakeholders, contractors, and teams is critical to effectively managing the complexities of megaprojects effectively (Aarons, 2006; Flyvberg, 2017).

Demonstrate Trust: Effective megaproject leaders create an environment that actively shows that they believe their team acts in good faith and will do their best to help the project achieve its goals. Being a reliable leader is being someone people can trust. You follow through on plans and keep your promises. You say something, and then you back it up with actions. Being dependable means meeting deadlines, being honest, and coming through in the clutch (Royse, 2021).

Resolve Conflict: Conflict resolution is the ability to peacefully settle disagreements between multiple parties. Effective conflict resolution is critical for managing diverse stakeholder interest and maintain project alignment (Flyvberg, 2017).

Coach and Mentor: Help to develop leadership skills, foster collaboration, and enhance trust and performance. Teaching and mentoring are about a leader thinking less about his or herself and more about how the team and overall company. When leaders teach and mentor, they help their teams develop professional skills and what they need to succeed at their jobs (Aarons, 2006).

Selecting and Developing Megaproject Leaders

Merrow and Nandurdikar highlight the current problems with selecting megaproject leaders (Merrow & Nandurdikar, 2018). Most specifically, success on smaller-scale simpler projects is not a good predictor for success on large-scale complex projects. However, most organizations select megaproject leaders based on their performance on considerably smaller projects. How can we improve the selection of individuals for these critical roles?

First, it is critical that megaproject leaders have experience in the primary domain(s) and discipline(s) for the megaproject. Merrow and Nandurdikar highlight that the majority of effective megaproject leaders in their study had an engineering background (Merrow & Nandurdikar, 2018). Even those who had been program managers for several years often had at least an undergraduate education in a more technical area. Selecting for individuals who have depth in an area and breadth across others (a "T-shaped" professional) again will improve the probability that the individual will have the necessary skillset to successfully navigate megaprojects.

The "Big 5" personality traits of neuroticism, extraversion, openness, agreeableness, and conscientiousness can be a useful indicator of whether and individual has a personality appropriate to megaproject leadership (Roccas et al., 2002). In general, successful megaproject leaders score high in extraversion, openness, agreeableness, and conscientiousness and low in neuroticism. They also develop strong emotional intelligence (Merrow & Nandurdikar, 2018). By utilizing an existing and widely used tool like the Five Factor model as well as emotional intelligence measurement indicators, project sponsors and owners can understand the personal characteristics of potential leaders. Candidates who are less likely to be effective can be weeded out more rapidly.

The HelixEMP employability skills provide a measurable foundation of these skills in earlier career development and can be used to diagnose potential trouble spots in the skills and mindsets of program managers that may not make good megaproject leaders (McDermott et al., 2021). Critical Helix characteristics relevant to megaprojects include strong technical foundations (technical depth) that are expanding to multi-disciplinary and trans-disciplinary abilities (trans-disciplinary means collaborating across disciplines to create new knowledge). These are applied in domains and megaproject leaders build their depth of domain experience by expanding to new roles and working with new collaborators. They also develop knowledge in systems approaches that allow them to make decisions more holistically, leading to less rule- and plan-following and more adaptability to address uncertainties with flexible approaches. HelixSE found two core mindsets that aid in managing uncertainties: comfort with uncertainty and paradox mindset. These are both also measurable (McDermott et al., 2021).

Self-Leadership and Learning capabilities are associated with a person's orientation to learning—how they respond to risk, uncertainty, and challenge and their ability to purposefully learn their way forwards to design, engage, fail, and learn and generate new knowledge which improves or transforms the job to be done (McDermott et al., 2021).

Situational leadership (where effective leaders adapt their style to each situation) is the foundational model for team leadership and collaboration—associated with individual and group learning capacity as it is manifested in relationships between people who are aligned around achieving a shared purpose of value (McDermott et al., 2021). This adaptation is essential in team learning. As a leadership and management career evolves, time as an executive leader—particularly a change leader—helps to increase core leadership competencies.

Megaproject leaders tend to have mastered transformational leadership (inspiration and followership). Doherty stated well the outcomes of this development in her study of the Heathrow Terminal 5 Project, what they called the "10/100/1000 rule":

> [Ten] senior leaders whose personal presence, vision and good judgement put the project on a course for success, often challenging existing industry norms. Another 100 leaders who by making critical differences, taking brave stands, interpreting new ideas and frameworks, leading by example, and ultimately creating an operating environment that enabled others to be successful. Another 1000 leaders who given that context were able to swim with the tide and do their

leadership role in a demanding workplace that had little space for error. (Doherty, 2008)

Conclusions

Effective leadership is the cornerstone of successful megaproject management, distinguishing projects that deliver on their promises from those that fail to meet cost, schedule, or outcome expectations. This paper has identified a comprehensive framework of six key competency areas—ranging from domain expertise and personal characteristics to advanced interpersonal and enterprise technical leadership skills—that are crucial for megaproject leaders. The findings underscore that while technical proficiency and experience in managing complex systems are essential, they must be complemented by strong interpersonal skills, a strategic mindset, and the ability to navigate uncertainty and risk. Moreover, the analysis reveals that successful megaproject leaders are characterized not only by their expertise but also by their capacity for self-awareness, openness, and a paradox mindset. Organizations should focus on these competencies when selecting and developing megaproject leaders, moving beyond conventional criteria that emphasize experience in smaller projects. Ultimately, by cultivating leaders with the right mix of skills and attributes, organizations can significantly increase the likelihood of megaproject success, achieving transformational impacts on infrastructure and society at large. This research offers a critical roadmap for understanding the complexities of megaproject leadership and suggests pathways for improving project outcomes through more effective leadership selection and development strategies.

Acknowledgements

This research was funded under contract HQ0034-19-D-0003 Research Task 0468 (WRT-1082). This research was conducted for the Defense Acquisition University. The authors would like to thank the sponsor for this work, Mr. David Gallop.

References

Aarons, G. A. (2006). Transformational and transactional leadership: association with attitudes toward evidence-based practice. *Psychiatr Serv. Aug, 57*(8),1162–9.

Doherty, S. (2008). *Heathrow's T5: History in the making*. John Wiley & Sons.

Eweje, J. Turner, R., & Müller, R. (2012). Maximizing strategic value from megaprojects: The influence of information-feed on decision-making by the project manager. *International Journal of Project Management, 30*(6), 639–651.

Flyvbjerg, B. (2017). *The Oxford handbook of megaproject management*. Oxford University Press.

Flyvbjerg, B. (2014). What you should know about megaprojects and why: An overview. *Project Management Journal, 45*(2), 6–19.

Gil, N., & Lundrigan, C. (2012, April). *The leadership and governance of megaprojects* (CID Technical Report No. 3/2012). Centre for Infrastructure Development (CID) and Manchester Business School, The University of Manchester.

Hutchison, N., Verma, D., Burke, P., See Tao, H. Y., Giffin, R., Yu, Z., Makwana, D., Chen, X., & Xiao, Y. (2020). *Atlas: Effective systems engineers and systems engineering* (SERC-2020-TR-007-A). Systems Engineering Research Center (SERC).

International Centre for Complex Project Management. (2021). *Complex project leadership—Competency standards* (Version 1.0). Edward Elgar Publishing Limited.

https://iccpm.com/resource-centre/complex-project-leadership-competency-standards-2023/

McDermott, T., Hutchison, N., & Crick, R. (2021). The evolution of HELIX: A competency model for complex problem solving. *Proceedings of the 31st Annual INCOSE International Symposium*.

McDermott, T., Yu, Z., Nadolski, M., & Clifford, M. (2024). The future of megaproject management. *Proceedings of the American Society for Engineering Management 2024 International Annual Conference*.

Merrow, E. (2011). *Industrial megaprojects*. John Wiley & Sons.

Merrow, E., & Nandurdikar, N. (2018). *Leading complex projects*. John Wiley & Sons.

National Institutes of Health. (n.d.). https://hr.nih.gov/working-nih/competencies/competencies-dictionary/political-savvy

Pennotti, M., Dominick, P., & Robinson, W. (2015). *RT-129: Advanced technical leadership* (SERC-2015-TR-101). Systems Engineering Research Center. https://sercproddata.s3.us-east-2.amazonaws.com/technical_reports/reports/1525983439-SERC-2015-TR-101_AdvancedTechnicalLeadership_RT129.pdf

Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The big five personality factors and personal values. *Personality and Social Psychology Bulletin 28*(6), 789–801. doi:10.1177/0146167202289008

Royse, M. (2021). 10 leadership skills every marketing leader should master. *Medium*.

Schoemaker, P. J. H., Krupp, S., & Howland, S. (2013). *Strategic leadership: The essential skills*. Havard Business Review.

Shenhar, A. J., & Dvir, D. (2007). *Reinventing project management: The diamond approach to successful growth and innovation*. Havard Business School Press.

Stevens, R. (2011). *Engineering mega-systems: The challenge of systems engineering in the information age*. CRC Press.

University of Pennsylvania. (2024). *Open-mindedness*. Authentic Happiness]. https://www.authentichappiness.sas.upenn.edu/newsletters/authentichappinesscoaching/open-mindedness#:~:text=Definition,of%20thinking%20for%20one's%20self

Van Wart, M. (2015). Evaluating transformational leaders: The challenging case of Eric Shinseki and the U.S. Department of Veterans Affairs. *Public Administration Review, 75*(5), 760–769.

THIS PAGE INTENTIONALLY LEFT BLANK

Defense Market(s): A Relook at the Explanatory Power of Several Economic Schools of Thought while Viewing the Defense Department as a Monopsony Firm

John D Driessnack—is a research faculty at the University of Maryland, Institute for Systems Research. He was the principal investigator at the University of Maryland, Project Management Center of Excellence for the OSD-funded Multidimensional Portfolio Management. OSD A&S funded that research. The research is currently being utilized within the Department of the Air Force. He is also a Professorial Lecturer at American University's Key Executive Leadership Program and Immediate Past President of the College of Performance Management. Driessnack owns Olde Stone Consulting LLC, a sole proprietorship. In recent years, he has been a member of two ANSI standard development teams and is currently on PMI's Standards Insight Committee. A retired USAF Lt Colonel, Driessnack's military experience includes six major defense programs, including senior PM for Global Broadcast System. Research and consulting interests include institutional economics and modeling, portfolio management, and quantitative performance management. [John.Driessnack@oldestoneconsulting.com]

Abstract

The paper proposes that the current emphasis on commercial practices ignores the challenges with price theory and competition and is a red herring for useful defense market analysis. Instead, we propose the use of New Institutional Economics, Public Choice, and the visible hand concept of the firm as a better economic model for assessing the market. Popular theory puts an emphasis on orchestrating a viable commercial market for defense products. This paper updates the premise that the defense unique market should not be compared to commercial markets and proposes a model that looks at the Defense Department as a monopsony firm within a complex government-influenced market. A useful economic model would balance improving efficiency by considering transaction costs within an aligned and integrated decision support system of institutions (requirements generation, resource allocations, and acquisition management) within the DoD as a firm. Creating an economic model is proposed using economic frameworks combined with current proposals to move the DoD from a program to an aligned portfolio management structure. A viable economic model should create a method to enhance understanding of how institutional changes affect the overall firm's performance in meeting its value chain strategy, given the market constraints.

Introduction

Concerns with the U.S. Military-Industrial Complex performance persist. The specific concerns, whether reaching back to Eisenhower's "unwarranted influence"[1] in the 1950s, or Operation Illwind's prosecution of 60-plus industry and government officials for procurement fraud in the 1980s,[2] or Gansler's "coordinated policies aimed at creating a viable market economy" in the 1980s and 1990s, or an example of the most recent concerns with the Center for Strategic & International Study on the U.S. Defense Industrial Base, "Isolated from the U.S. Economy," all have an overtone that the United States defense market is broken. The challenge is that we have no equal comparison to measure by, as our allies significantly rely on the U.S. defense institutional structures, and our adversaries utilize non-democratic institutional structures. We also have no economic models that are congruent with reality or meet what is generally accepted as properties of good models for the defense industry, especially the defense-unique items (Gabaiz, 2008). There appears to be no accepted set of economic

---

[1] https://www.archives.gov/milestone-documents/president-dwight-d-eisenhowers-farewell-address

[2] https://www.fbi.gov/history/famous-cases/operation-illwind

theories that have been put into an economic model for the Defense Market that is being used to guide effective and efficient changes in the institutional rules.

Over two decades ago, two economic schools of thought, New Institutional Economics and Public Choice, were proposed, which showed promise in illuminating what influences the Defense Market, providing policy insights, and could be used to build a good economic model (Driessnack, 2005). The institutions that govern the Military-Industrial Complex performance have continued to evolve over the past 60 years since the general framework of the current system was created in the 1960s. A significant change came with the consolidation of federal and defense contracting policy into a unified Federal Acquisition Regulations (FAR) in the 1980s, along with other centralization and consolidation efforts, such as creating a defense-wide plant and contract management with the Defense Contract Management Agency (Sapolsky, 1999).

Many policy efforts, including improving technology transition, seek free market mechanisms. Peck and Sherer, in the early 1960s, demonstrated that the defense market was significantly different from the commercial markets. Differences that are fundamental to the Defense Market institutions. Since the 1960s, even though there has been an almost continuous effort of reform, the sense is the U.S. Military-Industrial Complex is failing. Recent articles have titles noting the "crumbling foundation of America's military," noting the failed weapons and ammunition production to supply Ukraine. Others call for "immediate mobilization of a national industrial base capable of rapidly producing lethal, software-enabled hardware at scale" (*Strategic Edge: A Blueprint for Breakthroughs in Defense Innovation*, 2025).

Figure 1. Section 809 Panel Dynamic Marketplace Framework

The Constraints of a Dynamic Marketplace Framework

Recent articles and reports, such as *The Future Foundry, A New Strategy Approach to Military-Technical Advantage*, have interesting graphics which seem to imply a problem in the defense industry, such as the USAF Fighter Force Composition, and note that in the 1950s, the

Air Force employed 14 different fighter aircraft and employs only four today. It shows the production of hundreds of B-17s as if that is the goal. Technology today allows a couple of B-2s with advanced weapons to do the same, and more precise damage to our enemies with less loss of life on both sides. The report notes, "Clearly, this failure to change is not due to a lack of proposed solutions but is the consequence of inadequate political will and ineffective execution. Given broad acceptance among acquisition and industry professionals that the current system is flawed, endless recommendations for reforms, persistent bureaucratic intransigence, and a lack of meaningful change, how can the Department of Defense establish a reliable approach to generating and maintaining technological superiority in the 21st century?" (FitzGerald, 2016).

Palantir's paper, *The Defense Reformation*, notes the defense companies by market cap, noting that Palantir's Market cap is higher than any of the "traditional" hardware vendors. But they are a software company with revenue of $2.225 billion (CY23) with a market cap of more than $173 billion (Sankar (2024). It is in business that the Section 809 Panel would say is either Readily Available or Readily Available with Customization (see Figure 1). There are many industries that are in these categories, but they are not likely to be in with the unique defense systems, the platforms (unique aircraft, ships, tracked vehicles), nor the unique electronics, sensors, software, or weapons they carry. Comparison across the Section 809's Marketplace Framework should be done with caution.

A good example is Space X, with above $8 billion in revenue in CY23 and a market capitalization of more than $100 billion. Like Palantir, these are companies that are clearly participating in the commercial market and gain their market capitalization values from the potential commercial revenues in the future. They may be earning revenue and insight into technologies with their DoD and overall government contracts, but their valuations in the commercial market are from their potential reviews and profits within the civilian markets worldwide. Using these companies as a comparison to the traditional defense-unique companies can be very misleading.

Fallacy of The LAST SUPPER

Before we address some of the challenges that need to be addressed, we need to address the "The Last Supper" fallacy. The fallacy is that the defense market is broken because of the consolidation that happened post a 1993 dinner for two dozen of the military's biggest contractors hosted by then–Defense Secretary Les Aspin. This is known as "The Last Supper." It even has its own Wikipedia page, noting that the recent end of the Cold War had raised calls for a peace dividend, and Perry (the Deputy SECDEF) warned the defense industry that it would need to consolidate to survive upcoming budget cuts. The number of prime defense contractors in the U.S. was projected to decline from 51 to five in the following years (see Figure 2; DoD, 2022). The figure shows the consolidation. But is this a problem? Today, Raytheon, or RTX, derives only 59% of its revenue from defense, while Boeing gets 44%. Overall, the top five defense contractors together account for less than 20% of the defense budget. In comparison, in the broad retail sector, three companies—Walmart, Amazon, and CVS—command 50% of the market. Senator Roger Wicker, who introduced the Fostering Reform and Government Efficiency in Defense Act (FoRGED Act), wrote a companion article*, Restoring Freedom's Forge, American Innovation Unleashed* (Wicker, 2025). It references the "Last Supper" by referencing Palantir's *The Defense Reformation*, whose first section is named The Last Supper and Great Schism. The section starts by stating, "the most important consequence of the Last Supper wasn't a reduction in competition in the Defense Industrial Base, but the decoupling of commercial innovation from defense and the rise of the government monopoly." The underlying message is the implication that the consolidation of the defense industry and the lack of companies that do both defense and commercial business is an indication of a “great schism.”

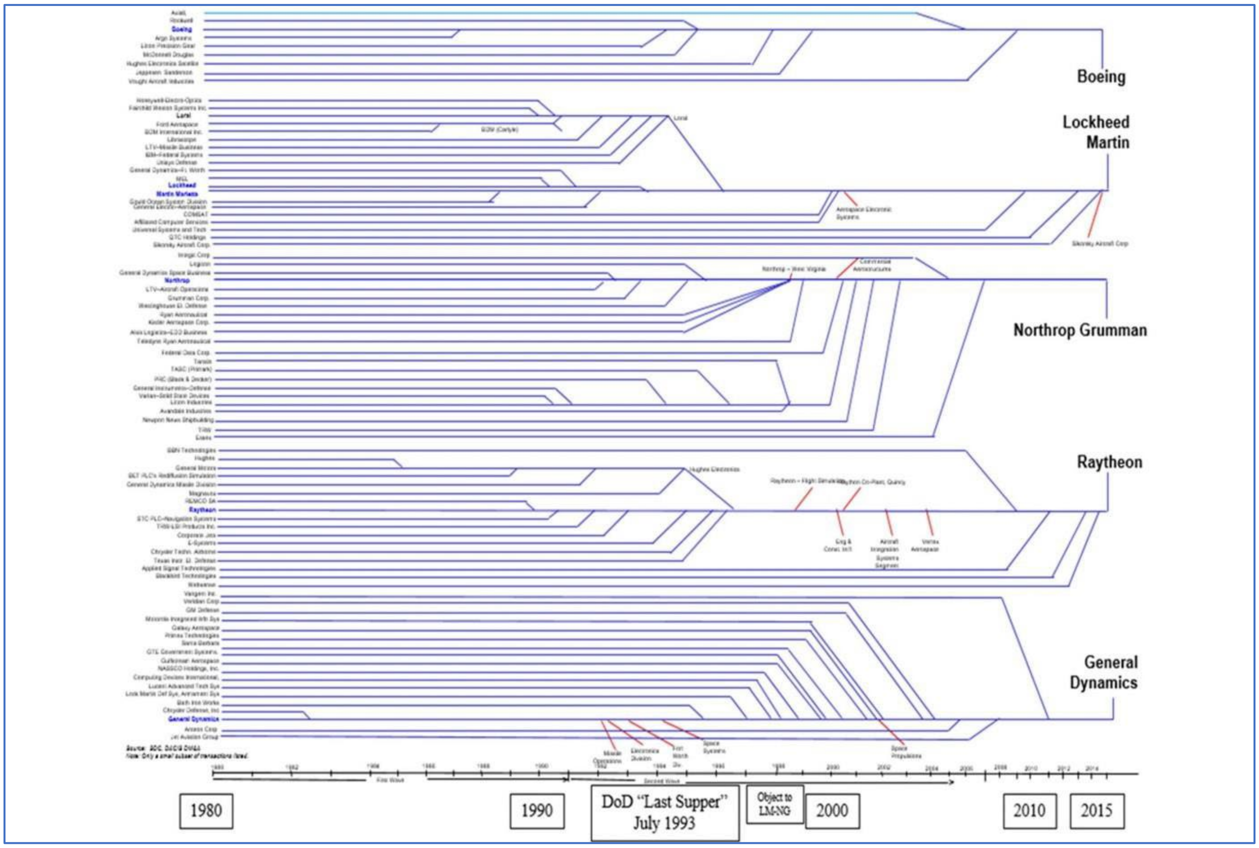


Figure 2. Defense Industry Consolidation

The Palantir paper states, “Before the fall of the Berlin Wall, only 6% of defense spending went to defense specialists—so-called traditionals. The vast majority of the spending went to companies that had both defense and commercial businesses.” The conjecture is that all those tedious regulations have made working in the national interest bad business. But is that really a problem? Most industries have consolidated, some, like Sears, which used to sell houses, are gone. Today, the top 10 companies get less than 20% of the total defense budget, and two of the top 10 are medical-related companies (Humana and Cencora). The landscape has changed, it doesn’t mean the market it broken. Look at the various global markets' share of each industry’s five largest firms. Defense doesn’t even make the list. Later in this paper, we will discuss how that small number of primes uses a vast market of subcontractors.

The National Defense Industrial Association (NDIA) reports the defense industry is made up of almost 60,000 companies with 1.1 million U.S. workers (Stewart et al., 2023). The 2021 number is a reduction of two-thirds from the 3 million workers in 1985. Again, this is reflective of other non-defense industries. The Industrial Production Index (see Figure 3) is higher in 2021 than is 1985, with more than half the labor removed! That production index doesn’t indicate an industry in trouble. The ups and downs reflect the volatility in the defense spending, which, if one followed “The Last Supper” theory when the post 9/11 Wars came why didn’t the number of prime contracts expand? The fact is, they kept consolidating.

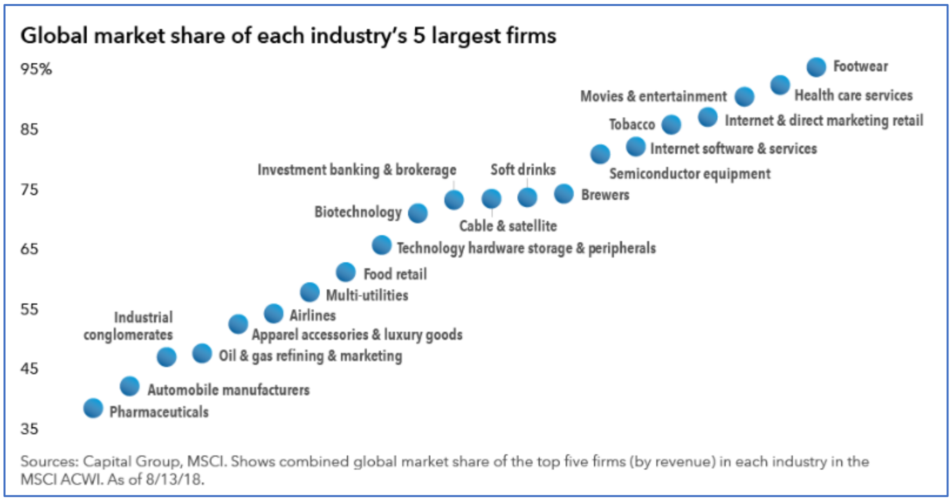


Figure 3. Global Market Share of Each Industry's Five Largest Firms

Industry Consolidation is Not the Problem

The Congressional Research Service, which publishes updates to its U.S. Defense Industrial Base (DIB) report (see Figure 4) notes up to World War II, the permanent DIB was operated by the government as arsenals and shipyards, and in "times of conflict, the armed services depended heavily on private contractors." Post World War II, "the wartime industrial base was not entirely dismantled, and the complex has been growing ever since. Government-operated facilities produced less than 10% of U.S. Defense equipment by the time President Eisenhower's speech on the military-industrial complex in 1961 (Congressional Research Service [CRS], 2024).[3]

Figure 4. U.S. Defense Production, 1947–2024

In the 1990s, the output of the commercial DIB decreased by approximately 35% and was sold as driving the infamous "Last Supper" in 1993. But as one can see in Figure 5, the next war, the war on terror increased expenditures, which decreased again with the Budget Control Act of 2011, but again, the shift to the great power competition has the expenditures growing again. But the "Last Supper" has not been reversed, which one would think should have happened, given the reasoning behind the "Last Supper" consolidation.

---

[3] The Industrial Production Index (IPI) is an economic indicator that measures the real output of the industrial sector in the United States.

Beyond the cyclical funding, the market is driven by a monopsony, highly regulated public sector, and the products are subject to restrictions on export and usage which are driven by conflict. The defense-unique part of the market is not a market based on competition and prices. Additionally, the DoD Small Business Strategy, January 2023, notes the number of direct awards from the DoD to more than 25,000 companies, which doesn't account for small businesses that work for larger prime vendors. It appears from this data that the market is robust. In Fiscal Year 2021, small businesses made up 73% of all companies that did business with the DoD, and 77% of the R&D companies. Moderna was a small business with DARPA and a grant recipient for MRNA vaccines. That turned out to be critical for the fight against COVID-19. A success story on numerous fronts relative to public policy and an indication how a small business can grow, as the company had over $7 billion in revenue in 2023.

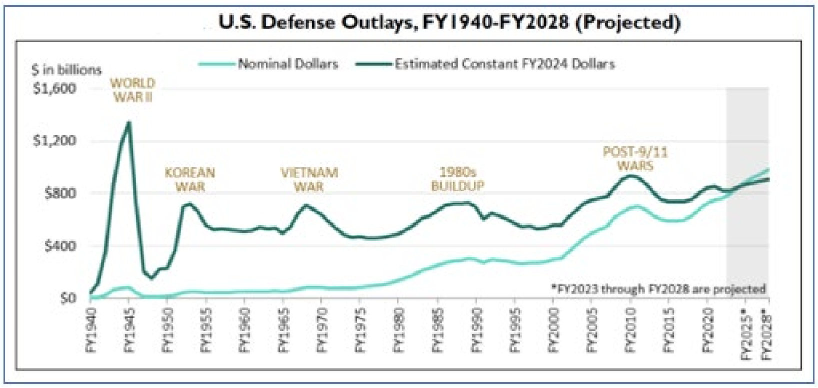


Figure 5. U.S. Defense Outlays

The other part of the defense-unique market is government-owned plants. Government-owned production and maintenance facilities, as well as ranges, test facilities, and federal-funded research and development centers (FFRDCs). Do we need more commercial or to go back to even more government-owned arsenals and shipyard models? The extent of these facilities is explained in the Congressional Research Service Report (CRS, 2024). Then there are the strategic National Stockpile and Petroleum Reserve. How do these fit into the calculation? The overall defense influence on many markets is complex and will not be solved by simply relying on commercial emphasis.

Another factor is the Defense Production Act of 1950, which allows the president to require private businesses to preferentially accept certain contracts and orders, as well as allocate materials, services, and facilities. It allows the president to provide loan guarantees, loans, purchases and purchase commitments, grants, and other financial assistance directly to private businesses (CRS, 2024). Additionally, the direction on defense needs appears in a law, called the National Defense Authorization Act (NDAA), which is often filled with thousands of changes each year. Not a commercial market situation.

“If the DIB is too small, it will be unable to supply all the materials, products, and services necessary to accomplish U.S. strategic objectives, and the military may lack the ability to execute its assigned missions. On the other hand, an industrial base with excess capacity could impose unnecessary financial costs on the U.S. government, requiring cuts to other programs, increased borrowing, or higher taxes. An oversized DIB may also distort the functioning of the country's market economy by diverting resources from other commercial applications”(CRS, 2024).

Market Below the Prime

If the major firms are doing their jobs, then the market below the prime contractor would be healthy. That does appear to be the case. It is not perfect, as with any market of subcontractors and vendors, but overall, with more than 60,000 vendors, it appears healthy. Again, there are some unique challenges within the defense market.

The DoD does need to be concerned with the industrial base beyond just the primes, with foreign intervention and overall stability. A proactive government policy toward high-technology industry mergers and acquisitions may be misguided due to the difficulty in predicting acquisition outcomes (King, 2003). A comprehensive F-22 Case Study looked at the concerns of the consolidation of the DIB and whether enough competition exists between remaining firms to maintain needed cost reduction and innovation. The study examined competition in the U.S. DIB by performing an in-depth case study of Lockheed Martin and the F-22 program that considers multiple tiers of the industrial base. The study found that defense firm specialization has led to outsourcing practices and arguably a more robust U.S. DIB (see Figures 6 and 7; King 2007).

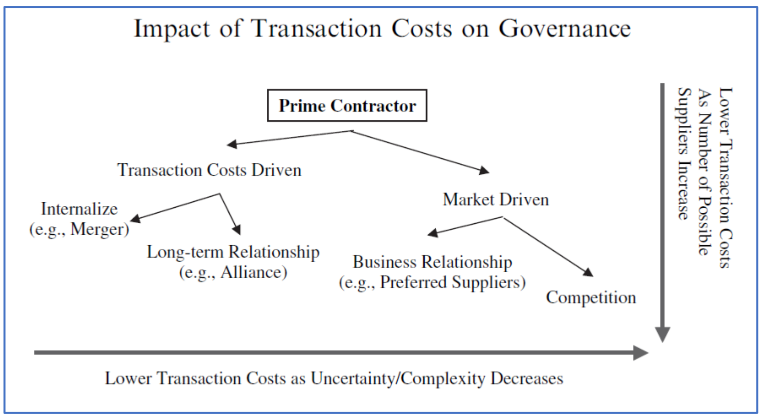


Figure 6. Prime Contractor

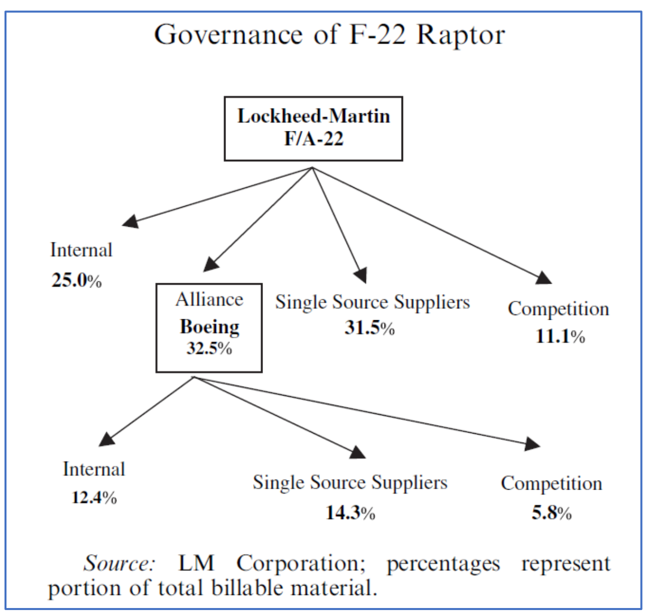


Figure 7. F-22 Prime Contractor

Below the prime contract, the transaction cost economics within the defense-unique market should hold with managers choosing the least costly method of organizing. Market exchange is generally considered more efficient than internalizing transactions, as it allows parties to a transaction to be competitively selected and drives the most efficient pricing for buyers and suppliers. However, Williamson suggests that market failure precludes market exchange and drives internalization of exchanges within a firm. Williamson originally outlined five situations that involve market failure:

1. **Bounded rationality**: Human beings tend to search for adequate and not optimum solutions.
2. **Uncertainty/complexity**: Conditions without readily discernible patterns or a manageable number of interactions that would facilitate decision making.
3. **Information impactedness**: Information asymmetry involving situations where one party is better informed than the other, making contractual arrangements difficult or expensive to verify.
4. **Opportunism**: Power imbalances that allow one party of a contractual relationship to pursue self-interests.
5. **Small numbers**: Reduced business choices resulting from limited quantities of either buyers or suppliers.

Later, a sixth market failure involving ‘‘**asset specificity**,’’ or a condition created from recurring transactions that create progressively stronger bilateral relationships, was identified.

What this outlines is why Oliver Williamson won his Nobel Prize in 2009. The theory is that transaction costs can drive decision makers to form long-term relationships, which are firms. This would be why some firms buy other firms, to lessen the transaction costs. Or what concerns some as industry consolidation.

Williamson uses the term “information-impactedness,” which refers to unequal access to information. Thus, Williamson takes issue with the idea that a firm is just another type of market. He emphasizes governance, as Alfred D. Chandler would in *The Visible Hand, The Managerial Revolution in American Business*.

Overall, the defense-unique market is highly influenced by national political considerations. This is where Public Choice Economics could potentially provide some insight. The federal and state governments provide guardrails for the commercial market to live within; those guardrails are on steroids when it comes to the national defense and the defense-unique market.

What Constraints Should be Considered?

The defense market has several unique characteristics, which pose unique constraints that need to be considered when creating an economic model to assist with understanding how changes will affect the market.

Constraint 1: Defense is Not a Single Market Sector

The defense market is not a uniform sector, but a diverse set of goods and services that are bought by the sovereign government, which often creates a monopsony (single buyer) environment in which commercial market forces, such as price theory, don’t work. The defense market is a microcosm of many markets that span almost every market type. It is hard to find a market where defense is not participating and often is a significant player. Each marketplace has different frameworks with different dynamics. They often interplay with each other, as outline by the Section 809 Panel. The 809 Panel created a simplified model of the dynamic marketplace framework for defense (see Figure 1). The “Readily Available” with or without “customization” is different than the “Defense-unique.” One needs to consider whether these frameworks should follow the same institutional rules.

With “Defense-unique,” the federal government is a monopsony buyer with exceptional control as the sovereign. The DoD even controls the foreign military sales, which is based on the fact that it is the sovereign and makes the rules and can change them at any time based on national security challenges. Plus, the market demands in this sector are determined by hostile actions of other governments and actors, for which it is almost impossible to predict how long they will endure the losses of materiel systems and the consumption of consumables. We need to be clear when talking about the Defense Marketplace, which framework we are discussing. For this paper, the focus is on the defense-unique framework. The other two markets can significantly use commercial best practices, and in many cases do today. The Readily Available with Customization can fall into either of the camps.

Constraint 2: Defense Needs to Meet the Public Policy Objective

In the readily available, with or without customization, the DoD often subjects itself to the same contract rules as private parties. Although there are exceptions set forth in federal statutes, regulations, and the Constitution to meet public policy objectives, such as small business goals. The challenge is these rules often don’t allow the federal government to take full advantage of commercial practices. Due to its special status as the sovereign and considering the statutes and regulations that apply to government contracting, government agencies are not in a position to take full advantage of the practices of the private sector. For example, agencies generally may not award contracts based solely on consideration of a company’s prior performance or enter into long-term strategic agreements. One needs to recognize the public policy challenge, but should not see it as a flaw in the performance of the market. Adjustment can and should be made, and often they are when national security is involved. In many cases, options are available to Defense officials when mission requirements need to truly outweigh public policy objectives. The level of adjustment is a balancing act which in the U.S. democracy will only be solved within the political system.

Constraint 3: Lack of a Stable Price Mechanism

Generally, there is a relationship between stable prices and stable markets. Hayek notes “because all the details of the changes constantly affecting the conditions of demand and supply of the different commodities can never be fully known, or quickly enough be collected and disseminated, by any one center, what is required is some [thing] . . ., which automatically records all the relevant effects of individual actions, and whose indications are at the same time the resultant of, and the guide for, all the individual decisions. This is precisely what the price system does under competition and which no other system even promises to accomplish” (Hayek, 1944, p. 36). Signals from changing relative prices are undoubtedly crucial and beneficial to economic decision-making (Issing, 2020). But, as noted by Peck and Sherer, “still much of the public discussion of weapons acquisition problems proceeds as if the terms ‘competition,’ ‘price,’ ‘buying,’ and ‘seller’ had the meanings they do in a market system.” Peck and Sherer were talking about the defense-unique market framework.

The instability comes in the demand for defense-unique products. The war in Ukraine with Russia is an example of the volatility in the defense market. As Lawrence Freedman, a professor emeritus of war studies at King’s College London, said in a recent article, “Clearly, I did not make the big call, which would have been to join those who had been convinced for some time that a big war was about to start. I was becoming increasingly persuaded of its possibility, but it still seemed to be such a self-evidently stupid move that I assumed that Putin had better options." The Hamas attack on Israel is another example. These events are hard to predict if they will happen, and then harder to predict how long they will last or how they will proceed. Ukraine thwarted Russia's efforts to seize Kyiv, and instead of a quick engagement, the war, along with the consumption of equipment and consumables, greatly increased demand in the defense market (Eckel, 2023).

No equivalent commercial market sector is a good model for the defense-unique market. During the COVID pandemic, we saw many commercial market mechanisms, such as offshoring fail. The consequence drove people to buy industrial rolls of toilet paper for their house, as the situation was not predicted within the market. Within defense-unqiue, there is a much greater risk for such near-term market failures that need to be considered when modeling the market.

Constraint 4: Defense Sets Relatively Risky Program Baselines

The Government Accountability Office (GAO) has had DoD Weapons Systems Acquisition on the high-risk list (GAO-25-107743) since 1990. In 2025, the DoD areas were reported to have gotten worse. The report notes, "Legislation—such as acquisition reforms required by the National Defense Authorization Act for Fiscal Year 2024—has prompted the DoD to consider how to address structural barriers that impede its progress in making change, such as its requirements processes. Despite policies that provide increased flexibility, the DoD continues to struggle to rapidly deliver complex, software-driven weapon systems." Anybody familiar with weapons systems acquisition, the significant part of the defense-unique market, is not surprised that the area has been on the high-risk list since the report started.

Is DoD leadership and the acquisition workforce such poor managers, or is the nature of the market just risky? If we stop making inappropriate comparisons, maybe we should consider that the defense-unique market is very risky and one method to address that risk is to embrace it and manage expectations. One way risk is taken is to set program baselines, the cost and schedule, at a relatively high-risk level. The DoD Cost Estimating Guide calls for a confidence interval for cost and schedule but does not set a confidence level. It notes the estimator should have "documentation showing management's acceptance of the cost estimate including recommendations for changes, feedback, and the level of contingency reserves decided upon to reach a desired level of confidence?" The Air Force policy, noted in AFI65-508, sets budget at the "mean of the program cost estimate distribution (typically 55%–65% confidence level)." A level of confidence that statistically guarantees significant cost and schedule overruns across a significant number of programs, which is exactly what happens. Why are we surprised by the outcome and characterize it as failure of the system?

There is no lack of understanding of how to estimate. The DoD and GAO have had a consensus on cost estimating best practices since the GAO published its first guide on the practice in 2009. The Air Force invented back in the mid-1960s, which the DoD adopted across the organization in the early 1970s, what has become the industry-wide best practice for performance management on programs, which is documented today in an International Standards Organization (ISO-21508) and American National Standards Institute (ANSI/PMI-006-2019) standard on Earned Value Management. Studies have shown the performance management technique, when implemented, is very effective. Many other practices, such as product teams, used widely in industry, evolved within the defense-unique market. Could it just be that the baseline by which we measure the market is optimistic, and that optimism is reflected in the results of the market?

A simple way to look at this is that the DoD starts 15 programs when they can only confidently afford 10. However, the out-year demand signal is so poor and unstable that it is not optimal to pick the 10 up front, but it is optimal to pick 15, set risky baselines, actively manage the programs, and cancel those that do not work out. We don't know because we don't have a method of economically modeling the defense-unique market, and there is no comparison market.

The DoD as the Firm, the Key "Visible Hand" in the Defense Market

In *The Defense Industry,* when talking about the defense industry,Jacques Gansler suggested in the late 1970s "to attach all of these problems, the government must implement a set of coordinated policies aimed at creating a viable market economy in each sector of the defense industry." He was calling for a managed commercial approach, what Alfred Chandler coined in his 1977 book, *The Visible Hand, The Managerial Revolution in American Business,* the "Visible Hand." Gansler's focus was on turning to commercial firms, which has been an emphasis by many for the past 60 years. But what if we looked at the DoD as the firm and the institutional rules it uses within the decisions support system, the requirements, the resources, and the acquisitions institutions. How would it adapt its institutions? One could look at the Services and the 4th estate agencies as companies within a larger corporation. The Office of the Secretary of Defense (OSD) and Joint Staff (JS) would be the top corporate structure with a dozen or so individual companies following generally the same rules but allowed to make modifications for their unique sector of the business. Given that structure, we could start to model The Visible Hand with the goal to optimize value, just like every other firm that is formed to beat the Invisible Hand of the market.

Two decades ago, Driessnack outlined at NPS's 2nd Annual Acquisitions Research Symposium an "alternative approach using transaction-costs analysis and the explanatory power of the New Institutional Economics and Public Choice School" for the concerns with the U.S. Military-Industrial complex (Driessnack, 2005). Driessnack referenced the Harvard Weapons Acquisition Process Study by Peck and Scherer in the early 1960s, which outlines that problem with "the public discussion of weapons acquisition problems proceeds as if the terms "competitions," "price," "buyer," and "seller" had the meaning they do in a market system. Driessnack's prior articles (2003, 2004) outlined the challenges with a commercial market type approach within the weapons acquisition monopsony market when the government is the sole buyer. The weapons market also has a unique characteristic of extreme uncertainty in demand, which is well discussed Mark Bowden's recent article, "The Crumbling Foundation of America's Military" (Bowden, 2024), which captures a quote by D.r Bill LaPlante, then the Under Secretary of Defense for Acquisition and Sustainment, on the Ukraine conflict with Russia. "But the idea that we would be spending or sending to another country 2 million rounds of 155"—the howitzer shells—"I don't think was really thought through." And if someone had raised the possibility, the response would have been: "I don't see that scenario."

Dr. Jacques Gansler in his dissertation, *The Diminishing Economic and Strategic Viability of the U.S. Defense Industrial Base* (Gansler, 1978) outlines "the problem" as "The Department of Defense is a monopsonistic buyer with—in peacetime—a shrinking buying power. What is the 'best' form of its supporting industrial structure to get the maximum defense capability for the approved dollars, with the least adverse impact on the public, and what actions—if any—should it be taking toward achieving this form of the industrial base?" Gansler states, "the single criterion that should be use for optimization of the defense industry is managerial efficiency, but this optimization must be done with the very clear constraints of the following: allocative efficiency, surge capability, flexibility for downward demand, research and development advancement, minimum impact on the overall U.S. economy, labor stability." His solution came in seven categories, coordinated government policy, integrated civil and military business, creation of a real market at the subcontractor and supplier level, new international policies in the defense industrial area, integrated and improved production surge, making cost a major decision criteria for all actions, and finally make the institutional changes necessary to achieve all of the above(Gansler, 1978). He had an opportunity to make his idea work when he was Under Secretary of Defense for Acquisition, Technology, and Logistics from late 1997 to 2001. Some changes were made, but is this really the answer for all of the defense market? Most large corporations have specific companies and divisions within those companies that are tailoring their institutions to the needs of the market. Can the answer come from a top-down set of policies or is the answer a more flexible system that allows more tailoring within the larger DoD firm!

Gansler, in a 2012 study on *Fixed-Price Development Contracts*, seemed to have evolved his thinking, noting, "When it comes to major development programs, there may be a good reason that the DoD has come to rely more on cost-reimbursement (as opposed to fixed-price) contracts. Unlike other DoD programs, MDAPs are often associated with a high level of uncertainty." The report also notes "agency theory, transaction cost theory (TCT), and incomplete contract theory provide a basis for understanding the advantages and disadvantages of cost-reimbursement and fixed price contracts." Unlike other DoD programs, MDAPs are often associated with a high level of uncertainty and "flexibility with regard to costs, schedule, and performance should be built into a contract so that trade-offs can be made as development progresses" (Gansler, 2012).

A more effective approach might be to focus on the DoD as a firm with various components that are tailored to the market it serves. In many ways, this is exactly what happens with the Government Services Administration (GSA) doing general buys, and more specific contracting happening within a particular subcomponent of the government. The DoD could consider the research done within firms to examine a better approach.

Model to Optimize the Visible Hand in Defense-Unique Market

Economic Theory for the DoD as the Monopoly Firm of the Defense-Unique Market

Two Nobel Prize economists, Oliver Williamson and Douglas North, note the lack of usefulness of classic rational choice and frictionless efficient markets as an effective model. Williamson provides a concept when "an outcome for which no feasible superior alternative can be described and implemented with net gains is presumed to be efficient." Not to say that government institutions around weapons acquisition can't be improved, but to compare them with commercial markets with price theory is the wrong model and benchmark. North discusses "incremental change" while understanding "path dependence is the key to an analytical understanding of long-run economic change" (Driessnack, 2003). A useful model would incorporate presumed efficient ideas and start with the current system.

Dixit, in his book *The Making of Economic Policy, A Transaction-Cost Politics Perspective*, reviews the principal agent problems in government through the view of transaction costs (Driessnack, 2005). "The hypothetical ideal with observable efforts and Coasean bargaining between all principles and the agent would be the first best. Respecting the information asymmetry but allowing all principals to get together and offer a combined incentive scheme would give the second best. If the principals cannot be so united, their Nash equilibrium is, in general, a third-best. In these formal terms, the result above says that the third-best outcome that is achieved has very low-powered incentives" (Dixit, 1996). The "low-powered incentives," which often do exist in government, are not proof of the inefficiency of government, but a recognition of the unique market environment. Bottom line, when judging the performance of a democratic, politically driven system, what "appears prima facie inefficient is in fact a reasonable way of striking a balance between the various interests, or multiple principles, given the transaction constraints." An alternative is a less democratic approach, which can be seen in the markets driven by competition and price, but as noted, that does not exist, nor will it exist, in the unique defense market.

Other views to consider include polycentric political systems and Public Choice approaches. These theories note that the "overall efficiency of any one part of the political system must envision the impacts on an overall political cost curve—not the individual political cost curse in any one section of the political systems." The noted challenge was to "embrace a more complete analysis utilizing New Institutional and Public Choice tools in a manner in which we can gain explanatory capabilities" (Driessnack, 2005). This can be seen with the often-improved efficiency when defense programs are classified (think Lockheed Skunkworks) and/or politics necessitates a selfless approach, as with the urgent need for Mine-Resistant Ambush Protected (MRAP) vehicle effort during desert storm.

The Emphasis on Flexibility and Portfolio Management

The Middle Tier Acquisition (MTA) framework, first authorized by Section 804 of the FY 2016 NDAA, is an acquisition pathway that focuses on delivering capabilities within two to five years. It allows for a very flexible path forward by breaking down what might otherwise be a major program into parts. A portfolio of MTA programs could be equal to a major defense acquisition program (MDAP). As per the interim implementing guidance, MTA is a "merit-based process for the consideration of innovative technologies and new capabilities [prototyping] . . . or existing products and proven technologies [fielding]." The rapid prototyping element of MTA must achieve residual operational capability within five years. The rapid fielding element must achieve initial production within six months and complete fielding within five years. Additionally, Section 804(d) establishes a Rapid Prototyping Fund to support MTA projects. The Rapid Prototyping Fund will operate with the onset of full MTA authority. In the interim, DoD components are funding their MTA efforts (Section 809 Panel Report, Volume 3 Section 1). The question is, will this work? MTA is just coming up on five years in enough numbers to look at whether they are making a difference.

Some programs, like Air Force Air Battle Management Systems (ABMS) under the new integration PEO, C3BM, under Program Element PE 0604003, with a budget more than $4.5 billion, have been broken into a number of MTAs and Software Pathway programs. ABMS could have entered into Milestone B as an MDAP but has taken the encouraged alternative approach and is known effectively a portfolio of smaller programs. What we don't know is if this approach will improve outcomes. It will be hard to know, because the DoD does not measure the alternative pathways similar to how it measures major programs, nor are there comparisons even at the model level, say a formal pre-milestone cost estimate or Analysis of Alternatives (AoA) that compared the strategy and would allow DoD to theorize if one approach over the other as being more efficient and effective.

Portfolio management within the management of programs has gotten popular in the last decade, both in government and industry. The Section 809 Panel noted with defense-unique development, "Much of the traditional debate surrounding acquisition reform is focused on the systems within this segment, and many challenges remain. While the DoD can still improve policy and process, its fundamental structure is appropriate." The panel did recommend a significant structural change to the management, the use of Portfolio Management (Vol III, Section 2). A series of recommendations outlined the benefits of shifting the DoD from a program-centric execution model to a portfolio-based execution model (Section 809 Panel Report, Volume 3 Section 1).

The Packard Commission and the resulting Goldwater-Nichols Department of Defense Reorganization Act of 1986 created the Defense and Service Acquisition Executives (DAE and SAE) and the Program Executive Office (PEO) that oversaw the work of the Defense Program Manager. The goal was to create a streamlined management structure for the Program Manager. The structure was the start of a layered portfolio starting at the Defense level, through the Service to a PEO, which had a specific portfolio of programs. As noted, this is like a larger corporation breaking out a particular sector into a new firm which has various divisions.

The portfolio concept in the late 1980s was mainly focused on stock portfolios. The DoD use of PEOs created a focused portfolio structure before it was utilized in industry. In 2008, based on a *Best Practice: An Integrated Portfolio Management Approach to Weapon System Investment Could Improve DoD's Acquisition Outcomes* (GAO, 2007), the DoD Directive 7045.20, Capability Portfolio Management, was created. The same year, the American National Standards Institute issued the first Standard for Portfolio Management (ANSI/PMI-08-003-2008). The DoDD 7045.20 did not provide a new authority to the proposed Capability Portfolio Managers and thus did not have much of an impact. But it is worth noting that these portfolio approaches were almost 20 years post the move to PEOs.

In 2019, the Section 809 Panel recommended an Enterprise Portfolio Management framework that incorporates the most recent *2017 ANSI Standard for Portfolio Management* (4th edition) practices. In general, the recommendation is not implemented by the DoD, but the term the portfolio management has gotten popular within the DoD. At the DoD enterprise level, the various defense oversight department each has their own "portfolio" review process. Figure 8 outlines the various organizational portfolio reviews and the desire, by some, to synthesize and align the data structure across the reviews.

In 2024, the PPBE Commission also made recommendations for the DoD to take a portfolio approach, this time with the resourcing process. This was followed by Senator Wicker proposing in December 2024 the Fostering Reform and Government Efficiency in Defense (FoRGED) Act, which includes very similar proposals for portfolio management as the Section 809 Panel, to include the replacement of the PEO with a Portfolio Acquisition Executive (PAE). The name change to PAE recognized both the portfolio nature of the organization and the Acquisition Executive, the authority. The PAE could now be a kind of independent division within the firm, held accountable to a set of metrics.

Brian Shultz, a Defense Acquisition University (DAU) Professor, created a paradigm shift chart (see Figure 8) with seven noted shifts. The emphasis one gets out of this is the move to focus on warfighters' missions within an enterprise architecture.

Figure 1. Paradigm Shift

Current		Future
Program-Centric Acquisition		**Portfolio-Management Acquisition**
Cost, schedule, & performance metrics	→	Return on mission effectiveness metrics
Stable and complete requirements	→	Welcome changing and new requirements
Optimize individual program	→	Optimize mission capabilities
Long development & fielding cycles	→	Rapid fielding of mission capabilities
Lock-in funding and schedule	→	Shift resources to exploit new opportunities
Fragmented processes and stovepipes	→	Enterprise architectures across mission domains
Monolithic Kill Chains	→	Adaptive and unified kill networks

Source of figure: The author.

Figure 8. Program vs. Portfolio Management

Multidimensional Portfolio Management

Though progress has been made, the challenges to the organizational changes are daunting, and to this day, the portfolio approach is not well aligned with the DoD across the decision support systems. Unlike product/systems design and development that uses a common ontology, called the work breakdown structure (WBS) across cost, schedule, performance, contracts, and other product/systems level efforts, there is no such ontology for the portfolio level across the decision systems. In fact, the portfolio breakouts are not the same, and the DoD lacks a very good mapping of the breakouts across requirements, resources, or acquisitions.

In 2024, OSD A&S, R&E, and Joint Staff signed an MOA to attempt to better align their three portfolio reviews. The other players, CAPE and Comptroller, did not join this effort. See Figure 9, which outlines the various portfolio reviews. IAPR is Integrated Acquisition Portfolio Review. CPMR is Capability Portfolio Management Review. TMTR is Technology Modernization Transition Review. With alignment of challenges and reviews not going well, in 2022, the OSD asked the Acquisition Innovative Research Center to explore the challenges with the portfolio approach. That resulted in the concept of not just a multi-layer portfolio structure, as outlined in the Section 809 panel, but a multidimensional structure that could incorporate the various ongoing portfolio review dimensions into an aligned vertically and horizontally ontology.

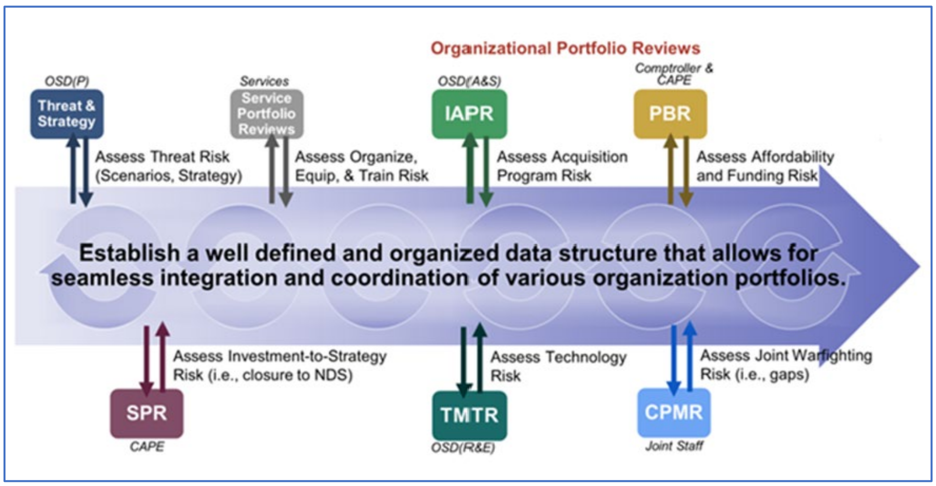


Figure 9. DoD Portfolio Reviews

The University of Maryland Project Management Center for Excellence led the study. They theorized the focus of the portfolio structure should align with a DoD value chain (see Figure10), which creates a horizontal path of primary activities that follow the defense-unique efforts from creating/capturing emerging technology through its incorporation into materiel systems vis programs of record. Then deploying those systems within operational units which combine various systems (platforms, weapons, sensors) to create a capability that is used within missions for a specific combatant, whether that be a regional, functional, or support combatant (UMD, 2023).

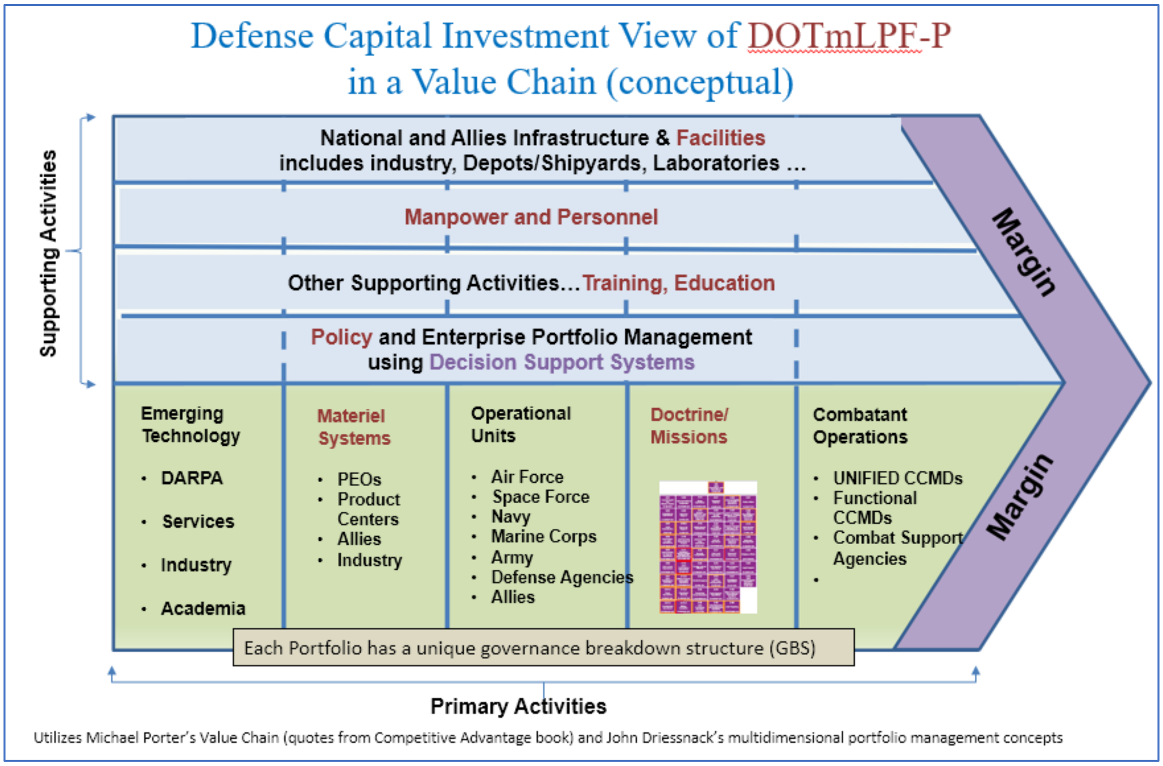


Figure 10. Defense Value Chain

Beyond the various multidimensional portfolio, the research proposed the need for an aligned ontology, which would require a many to many relationship, across the portfolios to understand the alignment of key information and decisions to be made within each of the portfolios. Given the time-critical nature of the defense enterprise, a robust network of schedule models that were also challenge-informed (considered constraints, assumptions, issues, risks, and opportunities [CAIRO]) was proposed. The only way to optimize across the enterprise was to have enough of an understanding on the alignment of the key primary activities. A notional model alignment and decision analysis model was proposed. See Figure 10, the Primary Activities are the multidimensional areas.

The research noted that the change would be a significant cultural change, and thus, implementation should be done with an agile approach in increments creating the tools necessary with minimum viable product (MVP) that continues to evolve over time. The approach would be like the big data efforts within industry, but industry systems would not meet the need. This unique firm, the DoD, needs unique ontology.

Conclusion

The DoD can't rely on a push to the commercial market will fix its challenges in the Defense-Unique Markets and in many cases the Readily Available with Customization Markets. The DoD does need to learn from industry and think of itself as a larger firm and design it's institutions to optimize across the various markets. One theme will not work within each of the markets it participates in, not across the diverse set of unique systems and products is invents, innovates, and managers. It should look at itself through the lens of a monopsony firm within the Defense-Unique Market and New Institutional Economics, with choices being assessed through the Theory of the Firm, Transaction Cost Analysis, and Public Choice Economics frameworks. Through this framework, further research and application are needed to evolve the institutions.

If one moves away from trying to be more commercial and looks internal to the DoD firm at the Decision Support Systems and the related organizations at the DoD (Enterprise) level, the Services, and then various multidimensional portfolios across the value chain, it is clear that the structure is not aligned, and thus one can conclude it is not streamlined. A move to a capability portfolio structure that is aligned across Requirements, Resources, and Acquisition (the three key parts of the decision support system) would be a great first step that is likely to have a significant return in effectiveness and efficiency. Given the complexity of the organizations and the institutional rules it must live by, further steps will be needed.

An economic model based on the appropriate theories, which are complex enough to be realistic but simple enough to be usable, will help with organizational change and the alignment of leadership and management. The DoD uses many models, including cost and schedule, missions and campaigns, manpower and force structure, and program and budgeting. However, there is no aligned model across the value chain, so there is no way to measure the impact of decisions against the department's key value measures across the enterprise.

This is not a simple task. Industry centers on price and profit, which allows it to simplify its modeling at the enterprise level and to optimize at the portfolio level. Price theory helps the commercial industry, which is not available to the DoD. Even though corporations, the firms, are managing with the “visible hand,” management is greatly assisted by the “invisible hand” of the market. The DoD, when it comes to the defense-unique market, has much less insight coming from the market thus the “invisible hand” will drive what will seem like extra transaction costs compared to industry. It is also much more influenced by “political transaction costs,” which are not as prevalent in most markets. Additionally, the DoD has a volatile market demand signal. Overall, it just might be more cost-effective to have more aggregate transaction costs than fewer.

The challenges in the defense market are unique, the solutions will need to be unique also, but they can be based on a combination of economic models and a selection of industry best practices along with defense unique solutions, which it has created before, and industry has adopted.

Acknowledgements

This paper is the first in a planned series under a funded (2023–2026) DoD Minerva Research Initiative (https://minerva.defense.gov/) at the University of Maryland Institute for Systems Research (ISR). The research title was ***Semantic Foundations and Formal Methods for Evolutionary Systems-of-Systems Architectures*** (https://minerva.defense.gov/Research/Funded-Projects/Article/3544550/evolutionary-system-of-system-architectures/). Semantic modeling was planned as a formal method to model key parts of the Defense Decisions Support Systems using New Institutional Economics and other economic frameworks. This research was cancelled by the Trump Administration just as the team was starting to frame the modeling efforts. This was likely do to the social science nature of the research and the current administration’s DOGE efforts.[4]

Refences

Bowden. (2024, December 17). The crumbling foundation of America’s military. *The Atlantic*.

Congressional Research Service. (2024, September) *The U.S. defense industrial base: Background and issues for Congress* (CRS R47751).

DoD. (2022, February). *State of competition within the defense industrial base*.

[4] https://www.fastcompany.com/91291503/defense-department-cuts-minerva-research-initiative-grants

Driessnack. (2003). A new institutional economics view of the consolidation of United States defense firms: An initial look. *Proceedings.* Budapest, Hungary: Institutions and Change 7th Annual Conference.

Driessnack & King. (2004). An initial look at technology and institutions on defense industry consolidation. *Acquisition Review Quarterly, 33*(1), 62–77.

Driessnack. (2004). US military-industrial complex transactional cost analysis integrated with public choice tenants: Air Force tanker lease case study. *Proceedings.* Vancouver, Canada: Western Economics Association International 79th Annual Conference.

Driessnack. (2005, May 18–19). *Unique transaction costs in defense market(s): The explanatory power of new institutional economics* (NPS-AM-05-024). Second Annual Acquisition Research Symposium of the Naval Postgraduate School.

Eckel, M. (2023, February 17). How did everybody get Ukraine invasion predictions so wrong?. *Radio Free Europe*. https://www.rferl.org/a/russia-ukraine-invasion-predictions-wrong-intelligence/32275740.html

FitzGerald. (2016). *Future foundry a new strategy approach to military-technology advantage*. Center for a New American Security. https://www.cnas.org/publications/reports/future-foundry

Gabaiz. (2008). *The seven properties of good models*. NYU Methodology Conference. https://scholar.harvard.edu/files/laibson/files/seven_properties_2008.pdf

Gansler. (1978). The diminishing economic and strategic biability of the US defense industrial base [Dissertation, The American University].

Gansler. (2012). *Fixed-priced development contracts: A historical perspective*. Center for Public Policy and Private Enterprise, School of Public Policy. https://spp.umd.edu/research-impact/publications/fixed-price-development-contracts-historical-perspective

GAO. (2007, March 20). *Best practice: An integrated portfolio management approach to weapon system investment could improve DoD's acquisition outcomes* (GAO-07-338).

Hayek. (1944). *The road to serfdom*. Routledge.

Issing, O. (2022, December 5). Why stable prices and stable markets are important and how they fit together https://www.ecb.europa.eu/press/key/date/2002/html/sp021205.en.html

King. (2007). Analysis of competition in the defense industrial base: An F-22 case study. *Contemporary Economic Policy.*

King. (2003). Investigating the integration of acquired firms in high-technology industries: Implications for industrial policy. *Acquisition Review Quarterly.*

Peck, & Scherer. (1962). *The weapons acquisition process: An economic analysis*. Harvard University, Graduate School of Business.

Sankar. (2024). *The defense reformation*. Palantir. https://www.18theses.com/

Sapolsky. (1999). The defense monopoly. *Regulations 22*(3). https://www.cato.org/sites/cato.org/files/serials/files/regulation/1999/10/defensemonopoly.pdf

Schultz. (2020). *A portfolio management based acquisition model?*. https://www.dau.edu/sites/default/files/Migrate/DATLFiles/Jan-Feb2020/Schultz.pdf

Stewart, J., et al. (2023, February). *Vital signs 2024: Posturing the U.S. defense industrial base*. National Defense Industry Association. https://www.ndia.org/policy/publications/vital-signs

Strategic edge: A blueprint for breakthroughs in defense innovation. (2025, January).

Wicker, R. (2025, February 15). *Senator Wicker announces Pentagon reform and innovation proposal*. https://www.wicker.senate.gov/2024/12/senator-wicker-announces-pentagon-reform-and-innovation-proposa

AI-Based DPCAP FAR/DFARS Change Support Tool

Jose Ramirez-Marquez—is an Associate Professor and Division Director of Enterprise Science and Engineering in the School of Systems and Enterprises at Stevens Institute of Technology. His research focuses on the development of mathematical models for the analysis and computation of system operational effectiveness research and explores the interplay between data visualization and analytical decision-making. He conducted funded research for industry and government and published over 100 refereed manuscripts in technical journals, book chapters, and industry reports. He holds degrees from Rutgers University in Industrial Engineering (PhD and MSc) and Statistics (MSc) and from Universidad Nacional Autónoma de México in Actuarial Science. [jmarquez@stevens.edu, 201-216-8003]

Joshua Gorman—is a Graduate Student at Stevens Institute of Technology with a Bachelor of Science in Computer Science. Previously, Josh garnered valuable experience as a research assistant, actively contributing to diverse application development projects ranging from meteor strike modeling to nuclear fallout predictions. He is continuing his education as the accelerated Computer Science Master's program, concentrating in cloud computing and computer networking. [jgorman4@stevens.edu, 201-216-8003]

Akram Amer—is an Undergraduate Senior at Stevens Institute of Technology majoring in Software Engineering with a minor in Computer Science. As a research assistant, he contributed to the DFARS/NDAA AI project by developing backend tools to extract and summarize legislative and regulatory language, including historical analysis of U.S. Code and National Defense Authorization Acts (NDAAs). He also worked on Department of Defense projects analyzing Justification Books, building AI-based tools to process and visualize DoD budget documents and legal codes, with a focus on backend infrastructure, NLP pipelines, and system integration. [aamer@stevens.edu, 201-216-8003]

Douglas J. Buettner—is the AIRC Deputy Chief Scientist at the Stevens Institute of Technology and is an Adjunct Assistant Professor for the University of Utah's Systems Engineering Program. His doctoral research at the University of Southern California's Viterbi School of Engineering examined issues with DoD software-acquisition strategies, identifying acquisition optimizations. His research is found in Acquisition Games and Software Defects: Dynamics of the Software Developer's Dilemma. Buettner has BS and MS degrees in Physics, with a minor in Atmospheric Science from Oregon State University, and his PhD from USC is in Astronautical Engineering. [dbuettne@stevens.edu, 201-216-8300]

Brian Mayer—is a Research Scientist at the Sanghani Center. Brian works with sponsors and collaborators to help apply Sanghani faculty research to solve government, corporate, and societal problems. Brian has helped deploy several long-term government-funded research projects to government sponsors. Brian's research interests are in the areas of decision support, forecasting, intelligence, and human-guided systems. Brian received an MS and BS in Industrial and Systems Engineering from Virginia Tech. [bmayer@cs.vt.edu, 540-231-5907]

Patrick Butler—is a Senior Research Associate at the Sanghani Center. The main thrust of his research focuses on the IARPA-funded EMBERS project that uses open-source indicators such as tweets, news, blog, weather, and so forth, to forecast population-level events such as civil unrest, elections, and epidemics. He graduated with a PhD in computer science from Virginia Tech in 2014, during which time he was awarded a SMART Fellowship. His dissertation was in the area of using knowledge discovery for intelligence analysis. [pabutler@vt.edu 571-858-3107]

Naren Ramakrishnan—is the Thomas L. Phillips Professor of Engineering at Virginia Tech and leads AI and machine learning for Virginia Tech's Innovation Campus. He directs the Sanghani Center for AI and Data Analytics and the Amazon–Virginia Tech Initiative in Efficient and Robust Machine Learning. Ramakrishnan's research interests span data science, forecasting, urban analytics, recommender systems, and computational epidemiology. He is a fellow of the Association for Computing Machinery (ACM), the American Association for the Advancement of Science (AAAS), and the Institute of Electrical and Electronics Engineers (IEEE). [naren@cs.vt.edu, 571-858-3331]

Bradley Freedman—is graduate research assistant at Virginia Tech. [bradleytfreedman@vt.edu, 540-231-5907]

Abstract

The Department of Defense's Defense Pricing, Contracting, and Acquisition Policy Contract Policy Directorate in the Office of the Assistant Secretary of Defense is responsible for periodic updates to the Federal Acquisition Regulation (FAR) and Defense FAR Supplement (DFARS) based on changes in the National Defense Authorization Act (NDAA), Small Business Administration rule changes, U.S. Department of Labor rule changes, or from executive orders. Reading through and assessing these documents for changes that require corresponding changes to acquisition regulations is labor-intensive. Further, when rule changes are proposed to the public for comments, reading and summarizing these public comments can range from straightforward to very labor-intensive.

In this paper, we report our initial research results to greatly improve the efficiency of analyzing the NDAA language for required updates of the FAR and DFARS, and issuance of memoranda and guidance using artificial intelligence, including large language models and advanced natural language processing techniques to provide an improvement in staff efficiency for these laborious tasks.

Keywords: Large Language Models (LLMs); Natural Language Processing (NLP); Department of Defense (DoD); National Defense Authorization Act (NDAA); Federal Acquisition Regulation (FAR); Defense FAR Supplement (DFARS)

Introduction

The Defense Pricing, Contracting, and Acquisition Policy (DPCAP) Contract Policy (CP) Directorate in the Office of the Assistant Secretary of Defense (OSD) is responsible for pricing and contracting policy matters across the Department of Defense (DoD). They execute statutes, executive orders, and policies through the timely update of the Federal Acquisition Regulation (FAR) and Defense FAR Supplement (DFARS) and issuance of memoranda and guidance. Fundamentally, they enable operations through business systems and standards.

The DoD DPCAP is responsible for periodic updates to the DFARS based on changes in the National Defense Authorization Act (NDAA), Small Business Administration (SBA) rule changes, U.S. Department of Labor (USDOL) rule changes or through executive orders (EOs). Reading through the changes made necessary by these multiple sources to complete required DFARS updates is labor-intensive for DPCAP staff. It also requires knowledge of all the rules in the FAR/DFARS to ensure that changes are made appropriately and references are made to the correct sections of the FAR/DFARS.

Artificial intelligence (AI) is a powerful tool that can accomplish many tasks and improve what humans can accomplish, but it has its limitations, so system development is a deliberate process that should be guided by policy and end use. Limitations can include bias, explainability, and trustworthiness (i.e. the well-known large language model [LLM] hallucination problem). Proper policy and implementation can limit bias, increase accuracy, and improve human effectiveness. When implementing AI solutions, it is important to understand these limitations and to create environments where AI systems and humans work in tandem to obtain the best results possible. For many tasks, the critical importance of human judgment means AI should serve as a complementary tool to improve human efficiency rather than as a standalone solution.

This project establishes a foundation for providing a cost-effective, scalable, semiautomated capability for managing regulatory policy updates, ensuring long-term efficiency and adaptability. This paper shares the process of learning the FAR/DFARS change process and identifying AI methods to make the task easier for DPCAP subject matter experts (SMEs) who are currently executing the tasks and provide additional support in identifying necessary changes. First, the research team worked with DPCAP SMEs to document all the steps of the process (Section 2). Then, a literature review was completed to review and discuss potential solutions (Section 3). Finally, AI methods were developed to automate certain tasks and assist the DPCAP team and were incorporated into an initial prototype to demonstrate those capabilities (Section 4). The team shares benefits, lessons learned, and future work in the conclusion (Section 5)

DPCAP FAR/DFARS Change Process

The first task is to identify change text of interest in the NDAA. Typically, SMEs read the NDAA line by line to identify text of interest. The NDAA is a lengthy document and can take quite a while to review to find all the text of interest, even using standard document search mechanisms. AI natural language processing (NLP) methods can automate all of this.

The second task is, given a text of interest snippet from the NDAA (from Step 1), to identify the locations in the FAR/DFARS that need to be edited. SMEs must rely on their knowledge or keyword searches to associate the change text to the FAR/DFARS sections.

The third task is to generate new or edited text for the FAR/DFARS given the text of interest from the NDAA (Step 1) and the text in the matching FAR/DFARS section (from Step 2). If the NDAA text affects the application of a prior rule, then that rule must be edited to comply. If there is no matching text currently in the FAR/DFARS, then new text must be generated. The proposed FAR/DFARS text must meet the requirements of the NDAA text.

The fourth task is to publish the proposed text for public comment. The proposed text is published in the Federal Register.

The fifth task is to review the comments from the fourth task. Comments are received and posted on www.regulations.gov. The comments are grouped, summarized, and posted on www.regulations.gov and in the Federal Register with the final rule.

The sixth and final task is to make any changes to the proposed FAR/DFARS based on comments received during the comment period.

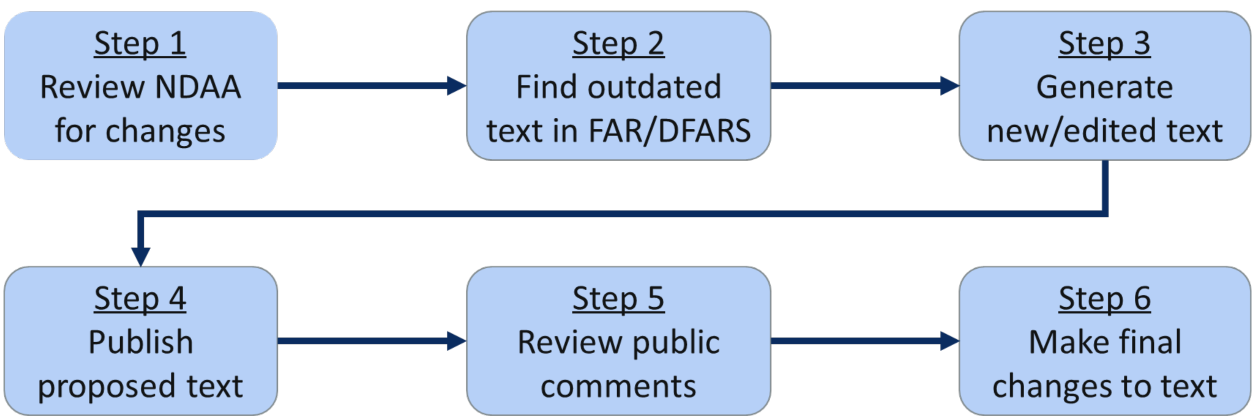


Figure 1. Task Flow Diagram

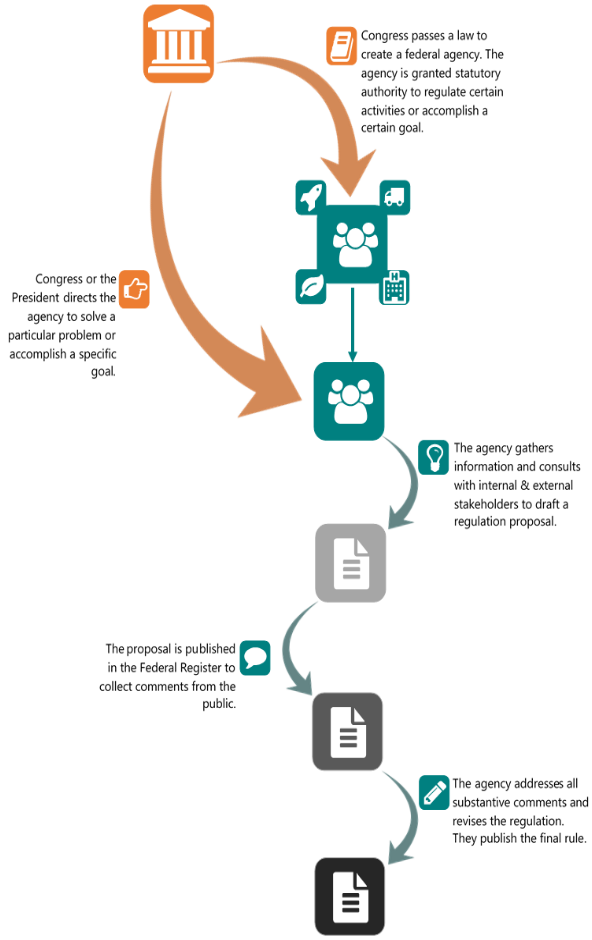


Figure 2. The Rule-making Process: How It Works (Public Comment Project, n.d.)

Potential AI Solutions and Limitations (Literature Review)

LLMs represent a class of machine learning (ML) systems trained on vast textual datasets that can comprehend and generate human-like text across a wide range of subjects and tasks. These models exhibit several key capabilities, including NLP and natural language (NL) generation with advanced contextual understanding. The research team investigated and discussed the adaptability of current LLM and AI approaches (Antón et al., 2023a, 2023b; Lewis et al., 2020; Neeser et al., 2024; Ramirez-Marquez et al., 2024; Zhang et al., 2024) that offered potential for automation and efficiency enhancement in the context of DFARS. The applications reviewed include automated information extraction, summary generation, query resolution, and the analysis of unstructured data.

Antón et al.'s (2023b) work was particularly relevant to the current project as it leveraged NLP and generative AI to enhance the identification of critical programs within DoD Comptroller Justification Books (J-Books) and improve the understanding of their budgetary implications. The work contained two phases, the first focused on utilizing NLP pattern matching to systematically extract and analyzing J-Book sections across different DoD branches, enabling the automated identification of key terms and their contextual significance, while the last incorporated analytics to aggregate data into portfolio budgets and integrated OpenAI's LLM to associate textual data with financial insights and visual analytics, ultimately enhancing decision-making and budget analysis.

Also relevant during the discussion was the work of Ramirez-Marquez et al. (2024), which explores the application of NLP techniques to enhance talent management and workforce adaptability within the DoD. Through analysis of text data from government, industry, and academic reports, NLP algorithms can automatically identify critical skills for the DoD workforce, particularly in acquisition and defense operations. The approach supports decision-makers by providing actionable insights to optimize talent acquisition, training, and resource allocation. This NLP-driven approach strengthens the DoD's ability to strategically develop and deploy personnel by automating skill identification that enhances workforce agility, reduces skill gaps, and improves operational readiness—critical factors in addressing evolving geopolitical challenges.

Fuzzy (also called approximate) string matching algorithms[1] are techniques used to compare and find similarities between text strings, even if they are not identical. These algorithms are useful when dealing with variations in spelling, typos, or slightly different wordings. In the context of LLMs, fuzzy string matching helps in

- **text preprocessing:** standardizing and normalizing text by identifying similar words or phrases.
- **information retrieval:** matching user queries with relevant documents, even if the wording differs.
- **entity recognition:** identifying names, locations, or terms that may appear in different forms.

When applied to LLMs, these techniques improve the model's ability to process and relate different text inputs, enhancing tasks such as document analysis, search functionality, and automated text generation.

Task 1: Identifying Text of Interest in the NDAA

One solution to this process is to identify keywords that typically indicate a FAR/DFARS change is necessary and use those keywords to find text of interest in the NDAA through simple searching. The limitation of this solution is that the keywords may be used many times not in a section that requires a FAR/DFARS change.

Another approach is to take ground truth examples (i.e., previous text that led to a FAR/DFARS change) and use AI to learn textual patterns that indicate a FAR/DFARS change. Document embeddings are a way to numerically represent documents of any length as vectors (Antón et al., 2023a, 2023b). The different sections or sentences can be compared to the ground truth examples using cosine similarity (distance between) or Jaccard similarity (Thada & Jaglan, 2013) to determine the similarity of the two vectors and the likelihood that the tested text indicates a FAR/DFARS change. The limitation of this solution is that the ground truth example text may be too broad and contain text that is not particularly indicative of a change. The embedding model may find similarities based on topic (e.g., acquisition system, country) versus impact. One way to mitigate this concern is to find many (hundreds/thousands) of ground truth examples and determine the features that are similar across the examples to determine what features should be looked for in the NDAA text segments. Another way to mitigate this problem is to use an LLM to identify similarities across ground truth examples which can be used to search for text of interest in the NDAA.

[1] See, for example, Approximate String Matching (n.d.).

A final suggested approach is to prompt LLMs to find text similar to the ground truth examples in a new NDAA using Retrieval Augmented Generation (RAG) to find references instead of generating text (Neeser et al., 2024).

Task 2: Identify Matching Text in the FAR/DFARS

It should be noted that the NDAA text may reference previous NDAA text (in a prior year) that led to a rule change. In this case, the FAR/DFARS section that resulted from the previous NDAA text is likely the section of the FAR/DFARS we want to identify.

One tool for accomplishing this is ElasticSearch (Lewis et al., 2020). ElasticSearch performs highly efficient keyword searching. The limitation of this solution is that words may be the same but will be used in a different context or for a different purpose.

Another approach is to use document embeddings similar to Task 1, but in this scenario, we can use LlamaIndex (Zhang et al., 2024) and Chroma vector stores, which are vector search solutions. These vectors are embeddings of the documents. A vector database is a collection of data that stores information as mathematical representations. We can search for vectors that are similar to the vector that represents the identified text of interest. This allows a more efficient search for multiple terms and more contextual information to be understood and returned (Schwaber-Cohen, 2024). The limitation of this solution is the same as it was for Task 1. One solution to this problem is to only search based on section titles, but this may not be specific enough.

If there is no matching text currently in the FAR/DFARS, then that likely means a new rule must be generated.

Task 3: Generate FAR/DFARS Text

If we have examples of known NDAA text (text of interest), matching FAR/DFARS text, and the newly proposed change to the FAR/DFARS text we can train a model on this information using *few-shot prompting* (Relevance AI, n.d.). The limitation of this solution is that few-shot prompting models may be overfit to the specific examples provided, leading to poor results when applied to new/different data. The success of few-shot prompting heavily relies on the quality and relevance of the examples provided. Poorly chosen or irrelevant examples can lead to inaccurate or nonsensical outputs.

Another approach is to use well-trained LLMs. LLMs are excellent tools for generating text. They can even generate text in specific styles and tones (Ullah, n.d.). If given the NDAA text of interest that indicates the new requirement, it could be asked to generate a rule or edit a current rule (if provided the current FAR/DFARS text). The limitation of this solution is that LLMs are liable to hallucinate. The FAR/DFARS rules provide specific information (e.g., numerical/section references), which are ripe for hallucination. There are a couple of ways to mitigate this. Numbering in FAR/DFARS is often a reference to other sections of that document and often those numbers may be new because the rules are new. New numbers could be provided to the LLM or entered after. Additionally, any text with references would be marked for review. Currently, SMEs use placeholders for these numbers when drafting the text manually. Also instead of asking the LLM to write the entire text we could ask the LLM just to rewrite the language that needs to be updated. The last potential mitigation is to fine-tune the LLM on a large set of government language documents or use RAG so it can learn these references better, effectively creating a reference library (Lewis et al., 2020).

To further enhance the output, we can ask the LLM to break down the prior rule before editing, similar to Chain of Thought (CoT) prompting (Wei et al., 2023). This helps the LLM perform more complex reasoning tasks by breaking down the problem into a series of intermediate steps. In this way, we are guiding the LLM to the solution instead of just asking for the output.

OpenAI's ChatGPT 4o can be used to generate summaries. ChatGPT 4o is representative of the capabilities of NIPRGPT, which is available internally to the DoD (Secretary of the Air Force Public Affairs, 2024). Llama (Grattafiori et al., 2024; Meta, n.d.) and Phi3 (Abdin et al., 2024; Microsoft/Phi-3CookBook, n.d.) are alternative open-source models that can be used. These models are small but still highly capable. They are deployed locally, which ensures complete control over sensitive data and documents, and are therefore compliant with government data protection regulations. Additionally, there is no reliance on external cloud services or data transfers. Ollama (n.d.) is also a framework that could help leverage various LLMs.

Task 5: Review Comments

Another task that LLMs excel at is summarization. The LLM will be prompted to simply summarize the comments (Zhang et al., 2023). Conveniently, the Regulations.gov API makes public comments to FAR/DFARS changes easily aggregable for input to an LLM. Document embeddings and clustering (Campello et al., 2013; Lloyd, 1982) can be used concurrently with LLMs to help structure and group the comments into similar topics and categories for more concise and usable summarizations. The limitation of this solution is that using LLMs incurs a cost, either time, resource, or monetary. The more data (text) you feed it, the more you use. Some proposed FAR/DFARS changes have thousands of comments. To mitigate the concern of costing too much, we propose narrowing down the number of comments fed into the LLM. We suggest doing this using the document embeddings and clustering results mentioned previously.

Task 6: Use Comments to Make Changes to the Proposed Text

This can be accomplished similarly to Task 3 but using the comments summary as an additional input and using the draft rule instead of the old rule. This will have similar limitations.

Proposed AI Solution Pipeline

We took the requirements from the SMEs and developed a prototype that demonstrates the potential to integrate the proposed techniques into the SME process in a way that enhances their effectiveness and efficiency while mitigating concerns due to limitations of the techniques and tools.

Task 1: Identifying Text of Interest in the NDAA

The research team worked on automating the process to identify language in the NDAA that could trigger regulatory changes. The team extracted relevant sections using key phrases used by the DPCAP staff to identify language that signals regulatory actions. The team mapped NDAA sections to historical DFARS rules through fuzzy string matching, ensuring accurate alignment. Additionally, the team developed a reusable, automated workflow for regulatory mapping, which can be extended to FAR and other regulatory frameworks for future applications. The team focused on

- **modification of existing tools:** Adapt existing AIRC NLP and LLM tools to ingest and analyze NDAA documents and historical FAR/DFARS language spanning 5 or more years.
- **keyword search and validation:** Develop AI-based algorithms to identify keywords and phrases in NDAA documents that signal potential DFARS changes. Validate identified changes by cross-referencing with historical documents.

- **change identification and suggestion:** Identify and track historical changes that correspond to DFARS and public comments.

To extract NDAA sections, NDAA data is read from Excel files containing sections of legislative text. A list of key phrases (e.g., “shall update regulations,” “modifies existing policy”) created with the help of SMEs is used to identify sections relevant to regulatory changes. Extracted sections are then stored with metadata, including source sheet and row index for traceability.

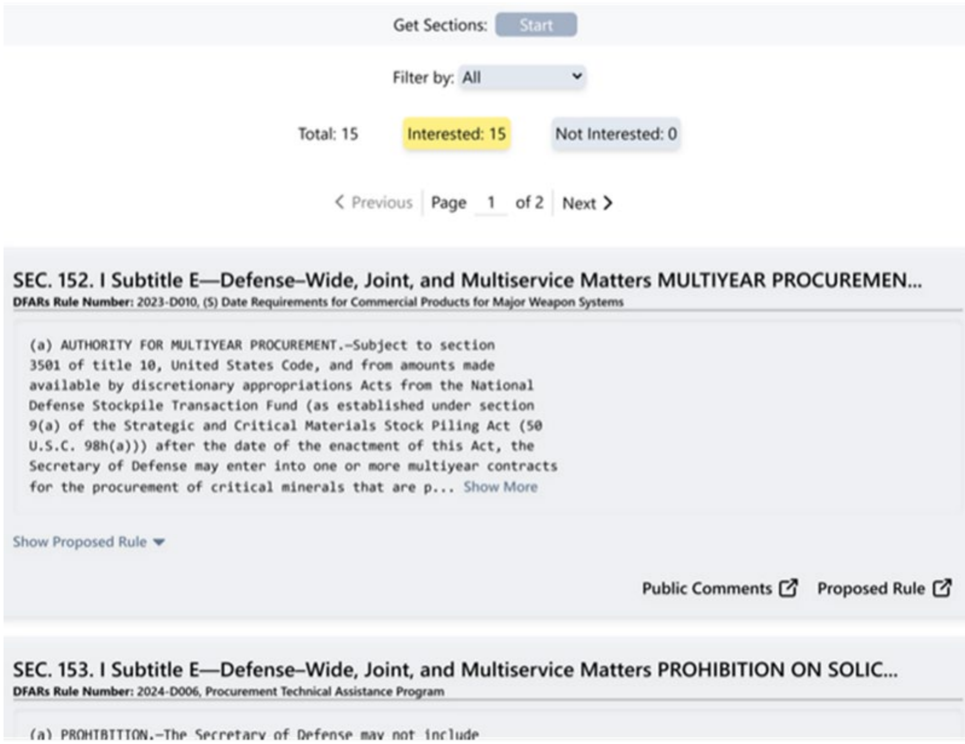


Figure 4. Screenshot of Demonstration Tool Extracting Potential Rule Change from an NDAA

Task 2: Identify Matching Text in the FAR/DFARS

To identify matching FAR/DFARS rules, first the rules must be extracted. DFARS rules are sourced from Word documents containing regulatory details. A parser scans document tables, extracting

- rule number (e.g., 252.225-7000)
- rule name
- comments link (if applicable)
- final rule notice

The extracted rules are then stored in a structured format for efficient lookup.

Once the rules are extracted, the NDAA sections can be mapped to these extracted rules. Both the NDAA and DFARS text undergo preprocessing to normalize wording and remove extraneous characters. Each NDAA section is compared to DFARS rules using fuzzy string matching (SequenceMatcher algorithm). The system selects the best-matching DFARS rule for each NDAA section based on similarity scores.

Figure 4. Screenshot of Demonstration Tool Providing Historical Proposed Rule Changes and Comments

Task 3: Generate (Draft) FAR/DFARS Text

We utilized some of the proposed techniques to create a tool that takes an NDAA year and section as well as a DFARS section as input (presumably as outputs from Task 2 and 3) and generates the text for a new rule and the corresponding text to be published with a proposed rule in the Federal Register (see Figures 5–7).

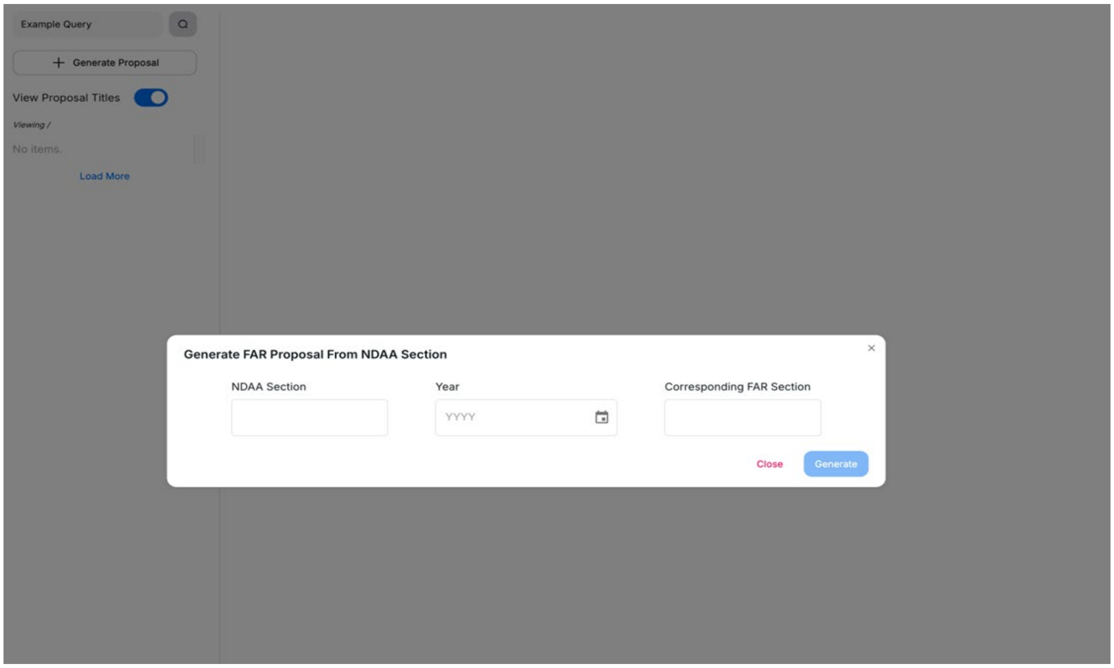


Figure 5. Inputting the NDAA Year and Section Along With the DFARS Section

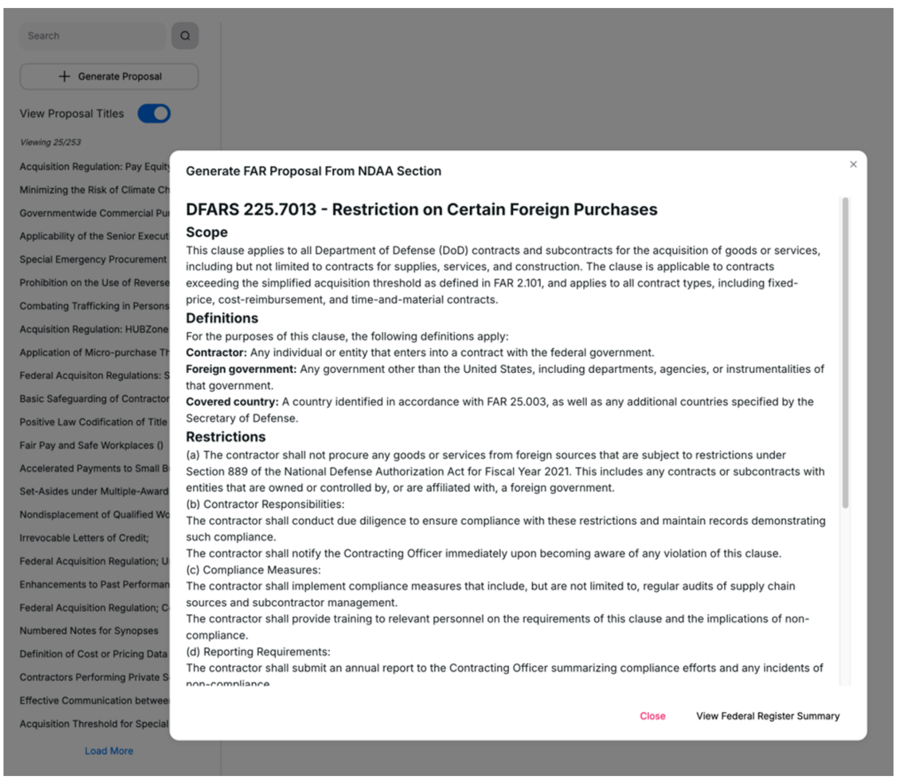


Figure 6. The Draft Proposed Rule Text Output by the Tool

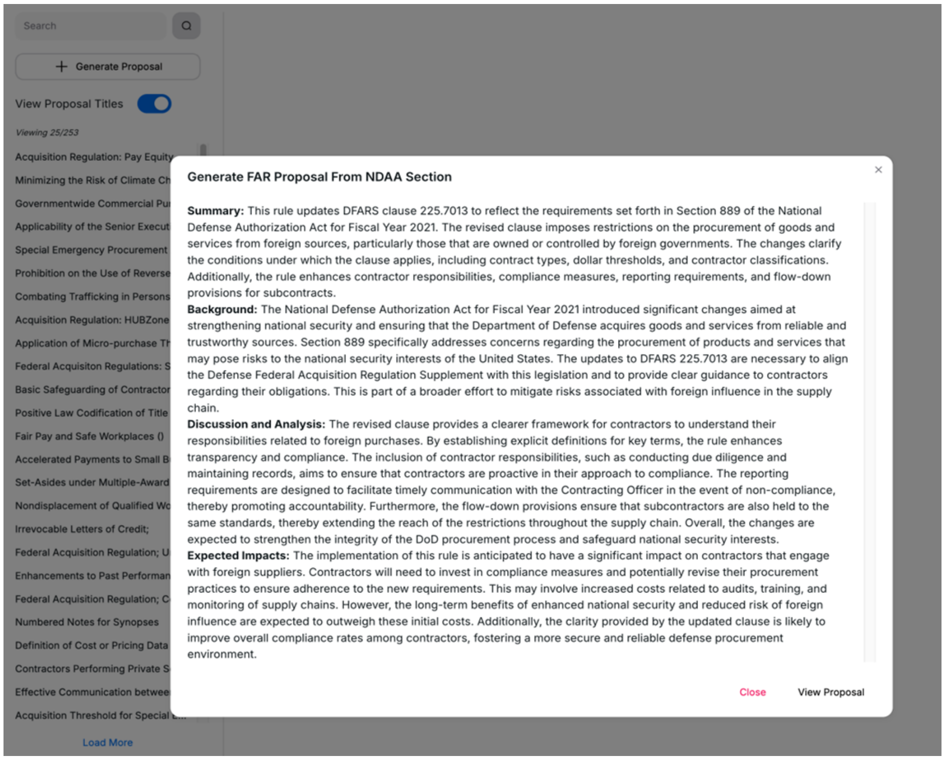


Figure 7. The Draft Text to be Published With the Proposed Rule in the Federal Register

The tool uses CoT prompting and requests specific steps and outputs to be performed to accomplish the task (see Figure 8). Virginia Tech researchers defined these steps. If SMEs provided a more accurate and thorough breakdown of the rule, the output could be improved through a more knowledgeable CoT process.

Maintain the same DFARS clause number and add or update sub clause numbers as necessary.

Maintain the same DFARS clause title.

Clarify under which conditions the clause applies, including specific contract types, dollar thresholds, or contractor classifications.

Maintain and amend the requirements from the current DFARS clause. Include contractor responsibilities, compliance measures, reporting requirements, and procedures.

Maintain references to other relevant FAR, DFARS, or government regulations and add any new references needed as a result of the NDAA text.

If and only if there are any flow-down provisions that specify when the clause must be included in subcontracts and, if so, under what conditions include or add them.

Provide a summary of the change, background of the change, discussion and analysis of the change and expected impacts of the change to be published in the Federal Register.

Figure 8. CoT Requests for Generating a Draft Rule and Corresponding Federal Register Text

Documents will be formatted similarly to the expected output but not exactly. The output of the tool (i.e., LLM) will need to be edited by an SME but can serve as a starting point. Reference and identification numbers will often have placeholders.

Note that as we test examples of the rule change, we are testing if the change already has a finalized rule that the tool will refer to for a new (or changed) rule. This is done instead of using the prior rule change and therefore impacts the output of the proposed text slightly. The research team tested the tool on both finalized rules and rules that were not yet finalized and expects to do more testing in a follow-on phase of this work.

Currently, the tool is only created and tested to generate DFARS text but should be easily extendable/applicable to creating FAR text.

Additionally, the tool currently requires a previous DFARS rule to be edited. If the NDAA requires a new rule to be created, the tool is not set up to generate that rule from scratch but could be modified (or improved) to accomplish this task as well. Tests were also not performed on the ability to generate a completely new rule. We tested our model and process on comments received from several proposed regulations. Example screenshots are shown in the figures that follow.

Task 5: Review Comments (Summarize)

We developed a prototype that can analyze, group, and summarize comments from any FAR/DFARS rule posted on www.regulations.gov. The user of the developed tool has the ability to select the proposed regulation they would like to review (see Figure 9), which is normally aligned on the left side of the page. The tool gets comments via the Regulations.gov API.

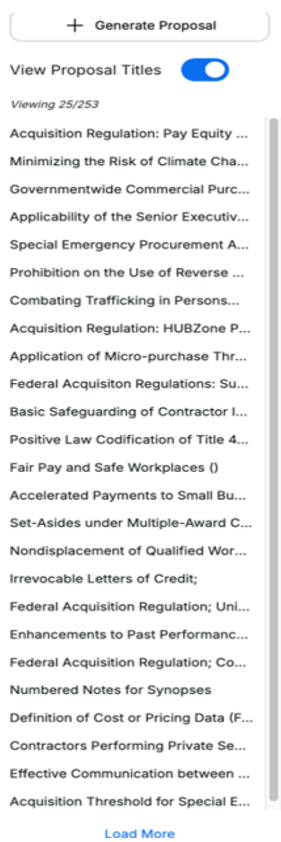


Figure 9. Users Can Select Any Proposed Rule Listed on Regulations.gov From the Column on the Left

Duplicate comments are removed, and similar comments (based on semantic meaning) are grouped in up to six groups. The groups are then summarized by an LLM (in this case GPT 4o; see Figure 10).

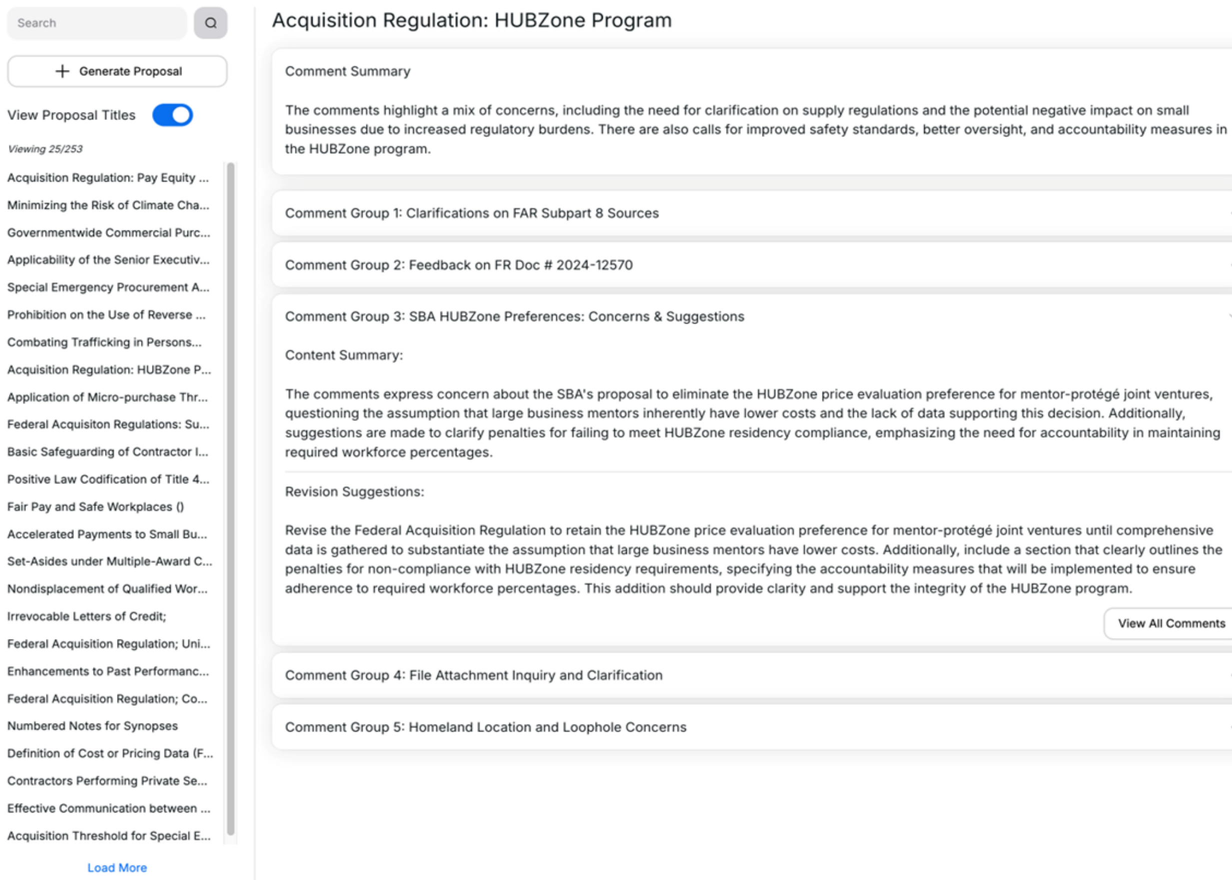


Figure 10. Groups of Similar Comments Are Summarized, and Suggestions for Changes Are Provided

If there are sufficient comments, a subset of the group's comments can be used for the summarization to reduce LLM token requests without drastically affecting the summary, since comments have been grouped based on similarity.

The interface of the tool allows the user to see all comments that are in a group (see Figure 11).

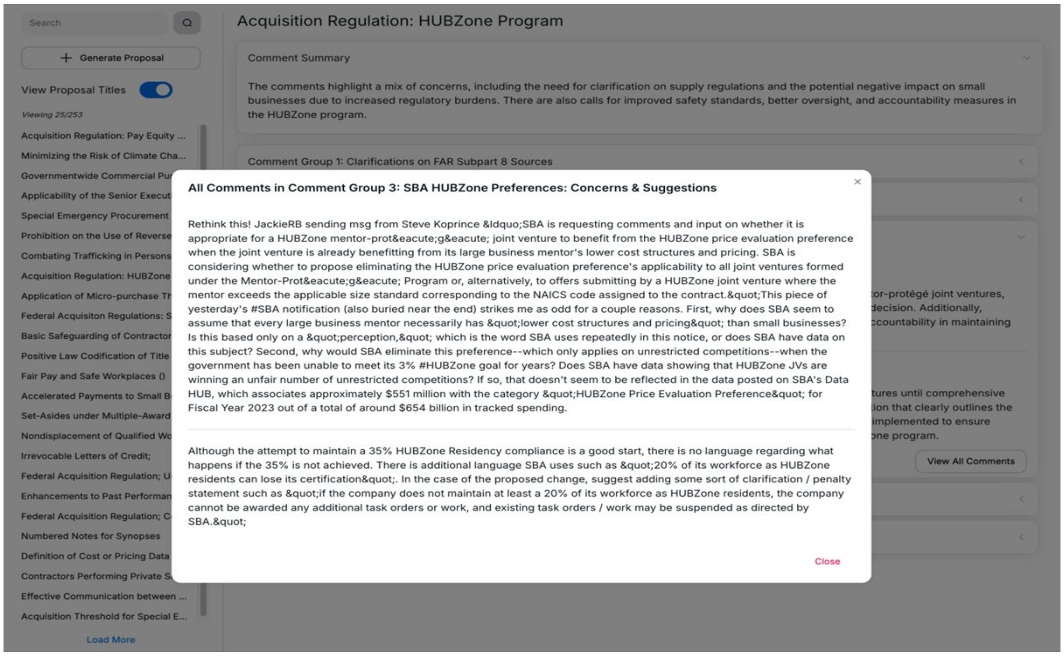


Figure 11. Actual Text of Comments Sorted by Group

Finally, the tool provides an overarching summary for the entire set of comments no matter how many groups and comments there are for the proposed rule (see Figure 12).

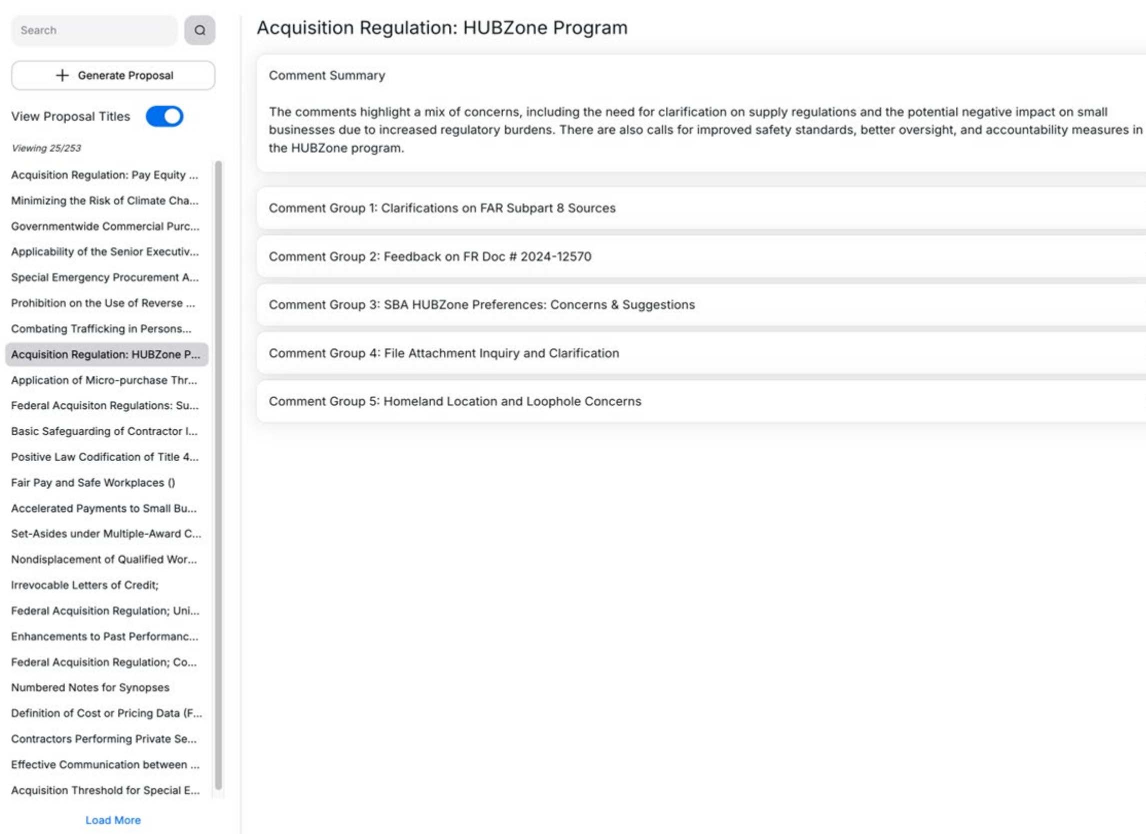


Figure 12. A Summary of All the Comments Regardless of Group

Task 6: Use Comments to Make Changes to the Proposed Text

For each group of comments, our tool provides edit recommendations to the rule based on what the comments are suggesting (see Figure 6). The user can choose whether to incorporate the suggestions from these comments or not.

The research team also tested providing the suggestions to the LLM along with the full text from the proposed rule (available through www.regulations.gov), but the text was often too long and more than the LLM token limits. There are solutions to this problem, but they were not addressed for this demonstration, since editing text was demonstrated for Task 3. Similar results (likely better because the text would actually be passed in) to Task 3 can be expected if an LLM were tasked to edit the proposed rule based on the comment suggestions shown.

Conclusions

Through the understanding that the research team gained about the FAR/DFARS change process from working with the DPCAP team, there is clearly an opportunity for efficiency and assistance through the careful implementation of AI techniques. Some techniques have limitations and therefore should be considered and implemented cautiously, but in many cases, some mitigations can be incorporated. The research team created an automated tool using AI to extract, analyze, and map NDAA sections to DFARS rules. These outputs are then fed to an LLM tool for generating proposed rule text and assessing public comments. The tool identifies key regulatory triggers and current changes, and streamlines the assessment of public comments on proposed rule changes. An executable web interface was delivered to DPCAP that integrates manual and automated steps, significantly reducing labor-intensive tasks for DPCAP staff.

The team would like to perform extensive testing of the tool on the NDAA implementation tracker for a specified period of years to validate its effectiveness and compare it to the current process. Additionally, updates can be made to improve performance and usability.

The prototype can be implemented immediately as is but has some limitations. We propose future development to address these limitations. Currently, the tool only finds and edits DFARS rules. The tool needs to be extended to work on the FAR as well. The LLM generation is relying on LLM knowledge of DFARS rules, but ingesting the raw FAR/DFARS text should enhance generated text. Another way to greatly improve generated text would be to have additional discussions with the SMEs to determine the structure and general requirements of FAR/DFARS rules. This would help inform the CoT prompting utilized by the LLM.

The current prototype is also limited to only suggesting edits to prior rules and only one rule at a time. Some rules require new DFARS sections/subsections, and in many cases, rules are updated together when an NDAA applies to multiple rules. A tool that matches this process would be more natural and easier to use for the DPCAP SMEs.

An easy implementation would be to allow SMEs to select which FAR/DFARS sections to update based on which NDAA rules, which would automatically feed the LLM rule generator (Step 3).

Finally, comments currently only result in the suggested changes, but a similar but more closed and refined process to Step 3 can be used to truly generate the final rule in Step 6.

Acknowledgements

We would like to thank Ms. Dawn Messer, OSD Defense Pricing, Contracting and Acquisition Policy, for her sponsorship and support of this analytical effort.

The views, findings, and conclusions in this document are solely those of the authors and do not necessarily reflect the views or positions of the U.S. government (including the DoD or any other government personnel) or the Stevens Institute of Technology.

References

Abdin, M., et al. (2024, August 30). *Phi-3 technical report: A highly capable language model locally on your phone*. ArXiv. https://www.doi.org/10.48550/arXiv.2404.14219

Antón, P. S., Buettner, D. J., & Ramirez-Marquez, J. E. (2023a). *ChatGPT budget justification book program element summarization: A proof-of-concept demonstration*. https://acqirc.org/wp-content/uploads/2024/05/03.-WRT-1079-AIRC-2023-TR-022-v2.0.pdf

Antón, P. S., Buettner, D. J., Ramirez-Marquez, J. E., Gorman, J. & Patel, A. (2023b). *DoD budget justification book keyword search association: A proof-of-concept demonstration*. https://acqirc.org/wp-content/uploads/2024/05/02.-WRT-1079-AIRC-2023-TR-021-v2.0.pdf

Approximate String Matching. (n.d.). In *Wikipedia*. Retrieved December 9, 2024, from https://en.wikipedia.org/wiki/Approximate_string_matching

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), *Advances in Knowledge Discovery and Data Mining* (pp. 160–172). Springer. https://www.doi.org/10.1007/978-3-642-37456-2_14

Cer, D., et al. (2018, April 12). *Universal sentence encoder*. ArXiv. https://www.doi.org/10.48550/arXiv.1803.11175

Elastic. (n.d.). *Elastic: The search AI company*. Retrieved December 9, 2024, from https://www.elastic.co/

Grattafiori, A., et al. (2024, November 23). *The Llama 3 herd of models*. ArXiv. https://www.doig.org/10.48550/arXiv.2407.21783

Le, Q. V., & Mikolov, T. (2014, May 22). *Distributed representations of sentences and documents*. ArXiv. https://www.doi.org/10.48550/arXiv.1405.4053

Lewis, P., et al. (2020, December). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (pp. 9459–9474).

LlamaIndex. (n.d.). *Using vector stores*. Retrieved February 6, 2025, from https://docs.llamaindex.ai/en/stable/community/integrations/vector_stores/

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Trans. Inf. Theory*, *28*(2), 129–137. https://www.doi.org10.1109/TIT.1982.1056489

Meta. (n.d.). *Introducing Meta Llama 3: The most capable openly available LLM to date*. Retrieved April 19, 2025, from https://ai.meta.com/blog/meta-llama-3/

Microsoft/Phi-3CookBook. (n.d.). *Jupyter notebook*. Retrieved December 9, 2024, from https://github.com/microsoft/Phi-3CookBook

Neeser, A., Butler, P., Mayer, B., & Ramakrishnan. (2024). *N. Bot automation using LLMs and plugins* [Unpublished manuscript].

Ollama. (n.d.). Retrieved December 9, 2024, from https://ollama.com

Public Comment Project. (n.d.). *How public comment works*. https://publiccommentproject.org/how-it-works

Ramirez-Marquez, J. E., Shafovaloff, G., Krzysko, M., & Verma, D. (2024, May). A semiautomated framework leveraging NLP for skill identification and talent management of the acquisition workforce in the Department of Defense. *Proceedings of the NPS Acquisition Research Symposium*.

Relevance AI. (n.d.). *Teach LLMs to mimic your style*. Retrieved December 9, 2024, from https://relevanceai.com/docs/example-use-cases/few-shot-prompting

Schwaber-Cohen, R. (2024, December 9). *What is a vector database & how does it work? Use cases + examples*. Pinecone. Retrieved December 9, 2024, from https://www.pinecone.io/learn/vector-database/

Secretary of the Air Force Public Affairs. (2024, June 10). *Department of the Air Force launches NIPRGPT*. https://www.af.mil/News/Article-Display/Article/3800809/department-of-the-air-force-launches-niprgpt/

Thada, V., & Jaglan, V. (2013, August). Comparison of Jaccard, Dice, Cosine similarity coefficient to find best fitness value for web retrieved documents using genetic algorithm. *International Journal of Innovations in Engineering and Technology*, *2*, 202–205.

Ullah, M. A. (n.d.). *Fine-tuning large language models to replicate scholarly writing styles: A step-by-step guide*. Medium. Retrieved December 9, 2024, from https://medium.com/@asadullah321/fine-tuning-large-language-models-to-replicate-scholarly-writing-styles-a-step-by-step-guide-945ac3a6d90f

Wei, J., et al. (2023, January 10). *Chain-of-thought prompting elicits reasoning in large language models*. ArXiv. https://www.doi.org/10.48550/arXiv.2201.11903

Zhang, T., Ladhak, F., Durmus, E., Liang, P., McKeown, K., & Hashimoto, T. B. (2023, January 31). *Benchmarking large language models for news summarization*. ArXiv. https://www.doig.org/10.48550/arXiv.2301.13848

Zhang, T., Ladhak, F., Durmus, E., Liang, P., McKeown, K., & Hashimoto, T. B. (2024, January). Benchmarking large language models for news summarization. *Transactions of the Association for Computational Linguistics*, *12*, 39–57. https://doi.org/10.1162/tacl_a_00632

THIS PAGE INTENTIONALLY LEFT BLANK

Supply Chain – Our Greatest Direct Cost

Kerry Meixner—is a Contract Price/Cost Analyst with DCMA's Cost and Pricing Command, where she has served for the past 5 years. She currently performs price and cost analyses on the various elements of contractor proposals and develops and recommends price/cost objectives for negotiations. Before this role, she served as a Contract Administrator with DCMA's Western Region. Prior to her federal government positions, she was an Assistant Vice President for 18 years in commercial banking. Meixner earned a Bachelor of Science degree in finance and risk management from California State University, Sacramento. She earned a Master of Science degree in business administration from the University of Phoenix. [kerry.k.meixner.civ@mail.mil]

Dayra J. Cole—is an Unlimited Contracting Officer with the Department of the Air Force and primarily has worked with the B-1 and B-2 Bomber programs. She currently is part of the Acquisition Exchange Program with Defense Pricing, Contracting, and Acquisition Policy. With experience supporting major Air Force programs, she has led high-value contract negotiations and strategic procurement efforts. Cole holds a Bachelor of Science in business administration from the University of California, Riverside. Prior to her federal government positions, she worked as a Staff Accountant for a Fire Life Safety construction company in Ontario, California. [dayra.j.cole.civ@mail.mil]

Peter Guinto—Mentor

Abstract

The Department of Defense (DoD) spends a lot of appropriated dollars and relies heavily on prime contractors to procure the goods and services needed to support our national defense. Prime contractors have been traditionally known as vertically integrated manufacturers, meaning they tightly control operations by taking ownership of many stages of the production process. However, there is now a perceived shift to horizontal integration, where they are relying more and more on subcontractors and suppliers for essential components and materials. If the perceived shift is significant enough, it raises important questions about transparency and the DoD's ability to effectively negotiate fair and reasonable contracts when they lack privity with key subcontractors. Through analysis of Cost and Software Data Reports and Defense Pricing, Contracting, and Acquisition Policy (DPCAP) Sole Source Peer Review data, this research confirms that direct material and subcontractor costs have increased as a percentage of total contract expenditures. While there may be ongoing efforts to optimize and potentially consolidate oversight functions within the DoD, addressing the challenges stemming from increasing reliance on subcontractors remains crucial. To address the identified challenges, policy recommendations will be made to include strengthening DoD oversight in a manner that complements ongoing optimization efforts, expanding the Industrial Base Analysis & Sustainment (IBAS) program, and implementing a Small Business Innovation Research (SBIR) Subcontractor Fast Track initiative.

Background

Although there is increased widespread perception that Department of Defense (DoD) subcontracting has grown over time, there has not been enough quantifiable research to review. In 1993, a dinner now known as the "Last Supper" at the Pentagon hosted by then-Secretary of Defense Les Aspin and his deputy, William Perry, was the catalyst of what was to become of the industrial defense base (Tirpak, 1998). The dinner was scheduled to serve as a notice that defense spending was going to fall rapidly. Following major cuts in defense budgets after the Cold War, the industrial base of defense contractors was forced to scale down operations or exit the market altogether. Of those that remained in the market, many were consolidated through mergers and acquisitions (M&A). As a result, the remaining competitors face having to be both suppliers and rivals in order to meet DoD demand. This dynamic known as "competimates" creates critical concerns when the subcontractors withhold crucial cost and pricing information from the prime due to proprietary concerns. In turn, government Contracting Officers are faced with ensuring fair pricing and effective negotiations when there is lack of information between competimates.

Since the 1990s, the defense sector has consolidated substantially, transitioning from 51 to only five aerospace and defense prime contractors. Therefore, the DoD is increasingly reliant on a small number of contractors for critical defense capabilities. For example, 90% of missiles procured come from only three sources. As a result, promoting competition and ensuring it is fair and open for future programs is a critical DoD priority (Office of the Under Secretary of Defense for Acquisition and Sustainment [OUSD(A&S)], 2022). The trend toward consolidation has continued over the past 8 years due to vertical and horizontal integrations and the entry of private equity firms performing roll ups. The increased consolidation within the defense industry can reduce the availability of key supplies and equipment, diminish vendors' incentives for innovation and performance in government contracts, and lead to supply chain vulnerabilities (OUSD[A&S], 2022).

The Issue

The DoD spends significant dollars on subcontracts and materials. In addition, as the industrial base has consolidated through M&A, competitors supplying each other (competimates) has become more and more common. In many of these instances, subcontractors do not share information critical to negotiations, information they deem proprietary, with the prime contractor. In these instances, government involvement where there is not privity of contract becomes critical. As we have determined through our research, the integration has increased over time, and it is in the DoD's best interest to adjust policies and procedures.

As the government works to modernize its weapon defense, it is increasingly procuring complex services and solutions that align with defense priorities. Updating military technologies and capabilities can lead to a higher proportion of costs to subcontractors and specialized expertise. With imminent emerging technologies, defense contractors have found vertical integration difficult to achieve when there is required specialized expertise across multiple fields and domains. In addition to the modernization of its defense base, the DoD also spends billions on the sustainment of its current fleets. As original equipment manufacturers (OEMs) discontinue parts, creating a pressing need to combat obsolescence and sustain critical defense weapon systems throughout their lifecycles, prime contractors are increasingly integrating legacy system sustainment into their portfolios to remain competitive.

While horizontal and vertical mergers affect the defense industrial base and present competition concerns, vertical mergers in particular have seen an increase in recent years. This vertical integration has the acquiring company controlling different stages of its supply chain, from raw materials to final production distribution. This integration and consolidation can help achieve efficiencies and reduce costs; however, it also is a concern for the DoD. For example, when a company has in-house capabilities down to the second and third-tier supplier levels, it can not only bid on a new platform as the prime contractor but as a "package deal," essentially selecting itself to provide subsystems. The problem with this is that other second and third-tier suppliers might never get a chance to bid on the subsystem work dominated by the prime, and the in-house division, facing no competitor, has little incentive to innovate or keep costs low. As time goes on, competitors may disappear from lack of work, innovation is further stifled, and prices go up (Tirpak, 1998). Ultimately, the DoD concern is that vertical mergers allow the buyer to take anticompetitive actions that provide an advantage over competitors.

The U.S. strategy of reorienting around great power competition in the 2010s and early 2020s has been an area of concern for DoD officials. In a 2022 report, the DoD found that consolidation had made it "increasingly reliant on a small number of contractors for critical defense capabilities" and observed that further "consolidations that reduce required capability and capacity and the depth of competition would have serious consequences for national security (Nicastro, 2024). One of the outcomes of consolidation, has been linked to overcharging by defense companies, due to the reduction of competition and suppliers has reduced government leverage in negotiating contracts.

A smaller defense base means frequent sole source environments in which prime contractors may have less incentive to aggressively control subcontractor costs. Not having a competitive environment means the contractors face less pressure to minimize overall costs. They might be willing to accept higher subcontractor costs if it simplifies their management. Similarly, the prime can simply pass subcontractor costs directly to the government, which potentially reduces their perceived need to scrutinize costs closely. As most sole source contracts are negotiated on the basis of cost plus a reasonable profit or fee, if a subcontract cost in the base year is higher, the future cost and therefore revenue/gross margin will grow. This significantly negates the prime contractor incentives to control subcontractor costs over time. The government needs to be vigilant in its oversight and use appropriate contract types and incentives to ensure cost control, even in the absence of competition in major program acquisitions.

Quantifying the Shift

This paper and research analyzed contract data from sources such as Defense Pricing, Contract and Acquisition Policy (DPCAP) Price Cost and Finance's (PCF) Sole Source Peer Review data throughout recent years, Government Accountability Office (GAO) assessments, Cost Assessment Data Enterprise's (CADE) Cost and Software Data Reports (CSDR), and insights from industry reports to contextualize these trends and identify areas in which there could be potential policy adjustments.

According to Defense Federal Acquisition Regulation Supplement (DFARS) 234.71, "all contracts, subcontracts, government-performed efforts, and major components (e.g., government furnished equipment), including FMS and programs in sustainment, regardless of acquisition phase and contract type, including non-FAR agreements, valued at more than $50 million, then-year dollars, for current and former ACAT I – II programs" are subject to cost data reporting requirements. As a result, contractors submit their reports to the CADE portal which were used for this analysis.

Cost and Software Data Reports Data

Two different data sets were compiled from CSDR. One report was for major Operations and Support (O&S) programs with data spanning from 2009–2025. The second report was for major Production programs with data from 2015–2025. Direct material costs, inclusive of subcontractor and vendor costs (and inter-work transfers) were compared against total direct costs to determine the proportion of material costs within overall contract expenditures for both data sets.

The below chart titled Comparison of Total Direct Material Subcontractor Costs Over Time presents a side-by-side comparison of the earliest and latest reported direct material and subcontractor cost percentages for various major defense programs providing O&S. The data revealed a consistent upward trend across most programs, indicating a definite increased reliance on subcontractor and material cost from prime contractors when comparing the earliest and latest available reports. While a small amount of programs only show a gradual increase, others—such as Programs 3, 5, and 18—show a dramatic shift. These upward increases support the hypothesis that prime defense contractors have fundamentally shifted their sourcing strategy and business models over time. This first data set consisted of 33 initial major programs which had available data through the CADE portal. However, of those, only 19 had enough data to evaluate.

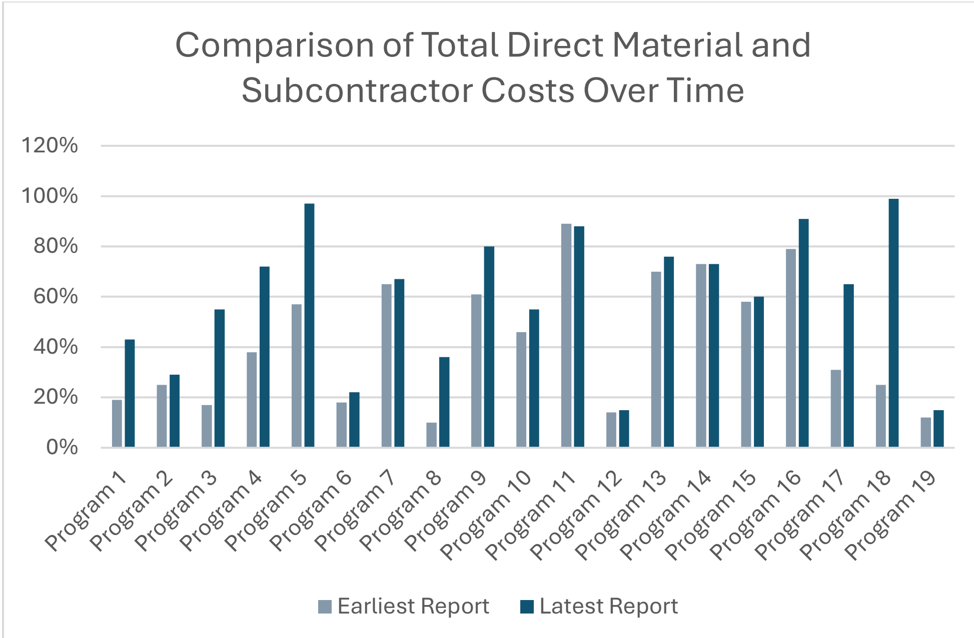


Figure 1. Comparison of Total Direct Material and Subcontractor Costs Over Time

Note: The data presented in this analysis was sourced from the Contractor Acquired Data Entry (CADE) portal, an internal database that is not publicly accessible.

The second data set, which analyzed the 2015-2025 major production programs, consisted of 67 programs with available data through the CADE portal. This data consistently showed a 71% average of total direct material and subcontractor costs. Additionally, it indicated slight overall percentage increases throughout the 10-year period for the majority of the programs. In summary, the following average percentages for 2015–2025 are as follows:

Calendar Years	Average Percentage
2015-2017	71%
2018-2020	73%
2021-2025	71%

Figure 2. Average Percentage for 2015–2025

Note: The data presented in this analysis was sourced from the Contractor Acquired Data Entry (CADE) portal, an internal database that is not publicly accessible.

The data analysis revealed a notable distinction between O&S and production programs. An examination of programs spanning 2009–2025 indicated that O&S programs exhibit consistently higher percentages of total direct material and subcontractor costs compared to production programs analyzed from 2015–2025. This disparity suggests that the unique demands of sustaining existing systems, particularly in the face of obsolescence and the diminishing availability of original equipment manufacturers (OEMs), may be driving prime contractors to rely more heavily on subcontracting for O&S activities.

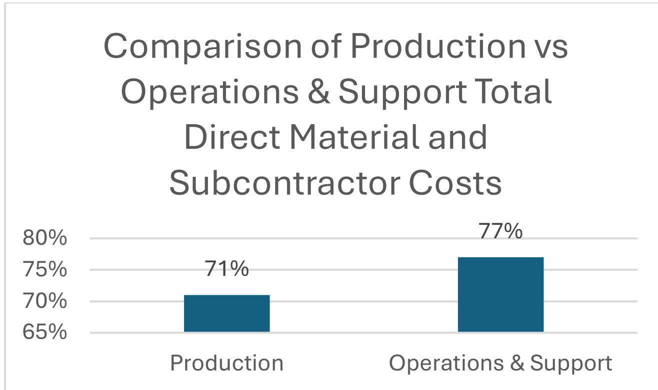


Figure 3. Comparison of Production vs. O&S Total Direct Material and Subcontractor Costs

Note: The data presented in this analysis was sourced from the Contractor Acquired Data Entry (CADE) portal, an internal database that is not publicly accessible.

While Cost and Software Data Reporting (CSDR) offers a wealth of data on defense program actual incurred costs, its potential is often underutilized within the acquisition community. Concerns exist that the data collected through CSDR is not consistently or effectively translated into actionable insights to inform acquisition decisions. This lack of rigorous follow-up limits the ability to leverage CSDR data for proactive cost management, contract negotiations and performance improvements throughout the program lifecycle. As a result, there is a need for enhanced mechanisms to ensure that CSDR findings are systematically integrated into the acquisition process.

Sole Source Peer Review Data

The quantitative analysis performed of DPCAP PCF's Sole Source Peer Reviews were based on data spanning from October 2023 to January 2025. This time encompassed almost 2.5 years of data focusing on preliminary and post negotiation memorandums from phase II of peer reviews from various military services such as Department of the Air Force (DAF), Army, Space Force, Missile Defense Agency (MDA), and many others. In a few rare occasions, preliminary negotiation memorandums from phase I peer reviews were utilized when post negotiation memorandums and phase II peer reviews were not available. The same formula from the CADE report was used to calculate the proportion of material costs within overall contract expenditures. This ratio was then averaged by fiscal year (FY) to reveal yearly trends in material cost allocation.

The FY 2022 data reflected an average percentage of 80.22%, which represented 14 non-service peer reviews. We excluded non-service contracts, such as Contractor Logistics Support, since these are more service oriented over a specific lifecycle. The FY 2023 data essentially mirrored the FY 2022 percentage, with 80.04%, which included 11 peer reviews. The FY 2024 data slightly went down as a percentage, to reflect 78.38%, which included 39 peer reviews. For FY 2025, we were only able to include three peer reviews that met our criteria (i.e., non-service), which had an average percentage for FY 2025 to date of 75.75%. As FY 2025 was only able to include three peer reviews, we will not include the resulting percentage in our analysis. In summary, the following average percentages for FY 2022–2024 are as follows:

Fiscal Year (October 1-September 30)	Average Annual Percentage
2022	80.22%
2023	80.04%
2024	78.38%

Figure 4. Average Annual Percentage per Fiscal Year

Note: The data presented in this analysis was sourced from DPCAP's Price Negotiation Memorandums (PNMs). This dataset is unpublished and not publicly accessible.

This data for the past 3 years reflects minimal change from each FY. We can conclude that the material as a percentage of direct costs have gone up the past 10 years, with the current data averaging about 78%–80%.

While rising direct costs certainly impact bottom lines, their ripple effects extend far beyond immediate budgetary concerns. One crucial area profoundly affected by these escalating costs is the defense supply chain, a complex network responsible for equipping and sustaining military forces. The increasing price of raw materials, manufacturing, and transportation creates significant challenges for maintaining a robust and responsive defense industrial base.

The Supply Chain

Supply chains rely on prime government contractors in order to function effectively. The United States' position as a leader in defense depends on a government supply chain that can keep up with the cost of demand (Greenwood Aerospace, 2023). Ideally, the DoD benefits from competitive market forces that form the basis for contract pricing, dictating the boundaries of what is fair and reasonable (Greenwood Aerospace, 2023). The ability to obtain data necessary to negotiate fair and reasonable prices has been particularly difficult for sole-source items.

The United States relies on its industrial base to provide and develop necessary technologies and weapon systems to maintain our national security objectives. Reliance on our industrial base poses risks, such as depending on foreign and single source suppliers and supply chain inefficiencies (GAO, 2022). This has created a challenge within the DoD, as suppliers for critical materials, such as replacing and upgrading obsolete parts on weapon systems, have not been immune to supply chain inefficiencies (over reliance on any single supplier??). Some of the factors that threaten the resilience of the defense supply chain are the declining capacity and competition in certain defense sectors (i.e., shipbuilding; GAO, 2022). This has caused a declination in the health of the defense industrial base, specifically with the DoD's supply chain, production capacity, and surge readiness, which are areas that are interconnected and are critical to U.S. national security interests.

The National Defense Authorization Act (NDAA) for FY 2025 contains provisions impacting government contractors and their supply chains. In particular, the NDAA requires the Secretary of Defense to implement policies, procedures, and tools to incentivize all DoD contractors to assess and monitor the entire DoD supply chain for potential vulnerabilities and noncompliance risks (Howard et al., 2024). If these vulnerabilities are not addressed within the defense industrial base, this leaves the nation exposed to supply chain disruptions and potential adversarial influence. In order to meet the mission of our armed forces, a healthy defense industrial base built on resiliency, diversity, and secure supply chains is essential (Shinego, 2024).

There has been a shift in subcontracts/materials in proportion to other direct costs. There should be a law/policy to provide a check against the government paying higher prices for contractors to cover their expenses to acquire companies in the supply chain, particularly where that business model precludes effective competition (Vergun, 2022). If nothing is done to combat this, these expenses will continue to be embedded in the contract prices taxpayers pay for products the warfighter must have to perform the mission. This ultimately will mean that the more we pay, the less combat capability we can acquire for a ready force (Vergun, 2022). The DoD needs to be able to perform adequate price reasonableness determinations. Therefore, changes such as legislative reforms are a necessity in order to ensure that the DoD stops paying excessive prices for essential parts/materials (Vergun, 2022).

Competition: The Driving Force

Competition is an indicator of the necessary industrial capability to deliver the systems, key technologies, services, materials, and products the DoD requires to support its mission. The DoD benefits from competitive markets via improved cost, schedule, and performance for products and services needed to support our national defense (OUSD[A&S], 2022). Incentivizing innovation through competition drives the defense industry to offer its best technical solutions at a best-value cost and price. Whereas insufficient competition may leave gaps in filling our mission needs, remove pressures to innovate to outpace other firms, result in higher costs to taxpayers as leading firms leverage their market position to charge more, and raise barriers for new entrants (OUSD[A&S], 2022).

A market that has many buyers and many sellers results in more competition, which drives the pricing for goods and services. The DoD aims to ensure that its contract obligations fund “fair and reasonable” contracts and do not allow contractors to gain excessive profits. The lack of competition may result in the types of excessive profits that the DoD aims to avoid (Congressional Research Service, 2023). When there are two or more offerors for a given contract, the DoD considers this as “adequate price competition.” This method incentivizes contractors to win the contract by bidding a lower price than their competitors.

Contractor Recommendations

Our research trend indicates that the material costs percentage is increasing as a percentage of direct costs. Is there anything that can be done to combat this trend? Is cost control a top priority for DoD contractors? It seems plausible that contractors overall would want to focus on controlling their costs in order to improve their profitability and drive long-term success. If contractors can effectively control and minimize costs, they can improve their competitiveness, increase profitability, and improve operational efficiency.

One of the ways that contractors can look at minimizing their costs is to see where they can reduce costs without reducing or compromising the product or service quality. Cost accounting can assist in helping contractors allocate expenses accurately, understand their cost structure, and make informed decisions regarding pricing, resource allocation, and project cost control strategies (Gowtham, 2024). Another tool that contractors can utilize is a project management technique called earned value management (EVM). EVM is a technique that helps companies monitor project costs, assess performance, and take corrective actions. It does this by integrating cost, schedule, and performance data. This assists in tracking the value of work completed in relation to the planned budget and schedule (Gowtham, 2024).

An additional step contractors can take to control costs is creating a financial plan or budget. This allows companies to effectively monitor their income and expenses over a specified time period. A budget will serve as a cost control measure by setting limits or targets for various cost categories. Monitoring actual expenses against the budget allows companies to identify deviations and take potential corrective actions (Gowtham, 2024). Effective budget control will help companies manage expenses and maintain financial discipline.

There are many ways in which contractors can apply strategies to assist in controlling costs. Some of these ways include supplier and inventory management, process optimization, waste reduction, and pricing strategies. One of the most important techniques to effectively manage costs is understanding inventory and supplier management. Contractors should continually focus on developing and maintaining strong relationships with their suppliers to negotiate favorable terms and conditions. This also includes maintaining clear communication, selecting reliable and cost-effective suppliers, and building collaborative partnerships to drive cost savings and improve overall supply chain efficiency (Gowtham, 2024). Effective inventory management monitors inventory levels to keep up with business demands, and controlling inventory to minimize carrying costs, reducing obsolete stock, in order to optimize cash flow. Companies can help control and minimize waste by implementing recycling programs, optimizing production processes to minimize scrap or rework, and promoting sustainable practices (Gowtham, 2024). This strategy aids in minimizing waste generation while maximizing resource utilization. Companies can use pricing strategies such as value-based pricing, cost-plus pricing, or dynamic pricing. By setting competitive prices that balance customer value and profitability, they achieve a better understanding of market dynamics, cost structure, in an overall effort to increase revenue (Gowtham, 2024).

While consolidation in the defense industrial base is looked at positively by many in order to gain efficiencies, these companies are thereby also decreasing the overall competition. The companies that vertically integrated can provide the larger defense companies the opportunity (if they choose) to potentially shut out as sellers those traditional second- and third-tier component suppliers who, operating at the lower end of the manufacturing “food chain,” normally sell to the “primes” (Tirpak, 1998). When consolidation occurs, it is important for the DoD to keep competition alive. As the supplier base narrows, it is important to have at least two sources in every sector to compete. Even in a sole-source environment, the DoD can offer ideas to keep competition and innovation alive. For example, a research and development effort can be started up for the next-generation system to create an alternative, rather than depend on one supplier. This also includes the prospect of dissimilar competition by having variants as an example of using different approaches to the mission itself (i.e., competing missiles versus airplanes).

Some argue that the consolidation will remove pressures to innovate and outpace other firms, and ultimately the taxpayer will suffer as leading defense contractors leverage their market position to charge more and raise barriers for new entrants (Tirpak, 2022). As an example, satellite suppliers have dwindled from eight to four over the past 30 plus years. Reduced competition and fewer suppliers will have an adverse effect in filling defense needs.

Subcontractor Management

Typically, prime contractors should have a strong incentive to manage and control costs at the subcontract level because they are financially responsible for the overall project budget and profit margin. In fact, FAR 42.202(e)(2) states that “the prime contractor is responsible for managing its subcontracts” (FAR, 2025). While prime contractors are accountable for delivering products or services within the agreed-upon budget, are they continually trying to monitor and control their subcontractor costs? One of the contract types, Firm Fixed Price (FFP), provides prime contractors with maximum incentive to control costs, as they are responsible for the risk of cost overruns. Prime contractors are ultimately responsible for the allowability of subcontract costs. When they are denied access to their subcontractor records, they need to request field pricing and evaluation of their proposals to determine fair and reasonable prices. Prime contractors need to effectively manage their subcontractors so they do not risk suboptimal program outcomes and risk being accused of excessive pass-through costs.

One of the ways to manage subcontractor costs is to correlate the most effective contract type with the product or service being procured. The risk associated with the work to be performed is an important factor when selecting a contract type. The contract type and the negotiated contract pricing are interrelated and, therefore, should be considered together. The contract type will include certain elements that create the contractor compensation arrangement. These will usually include any contract financing, profit or fee, and contract terms and conditions. In particular, the use of cost-plus-incentive-fee (CPIF) and fixed-price-incentive Firm Target (FPIF) contracts are highly correlated with programs that achieved better cost and schedule outcomes (OUSD[AT&L], 2016). Through incentives, such as CPIF, the prime contractor can earn more profit/fee by reducing cost, exceeding performance objectives or achieving the desired schedule. An incentive-type contract can allow the government to share in cost savings if structured appropriately. The profit/fee motive is what incentivizes the prime contractor by providing the opportunity to realize an increased profit for attaining cost, performance, or schedule criteria. Negative incentives can also be used to motivate contractors to avoid reduced profitability when desired outcomes fall short. Incentive contracts can be structured to achieve desired objectives through reasonable and attainable targets (OUSD[AT&L], 2016).

Policy Recommendations

Based on the findings which indicate a measurable shift in more direct materials and subcontractor expenditure within defense spending, it is important to acknowledge the impact on cost efficiency, transparency, and DoD Contracting Officers' ability to ensure fair and reasonable pricing. Given the aforementioned challenges, it is imperative there be policy adjustments to strengthen oversight and transparency throughout the DoD.

Increase the DoD's Oversight Capabilities

One of the key challenges that Contracting Officers repeatedly encounter is the inability of prime contractors to obtain cost or pricing data from their subcontractors due to proprietary data resulting in excessive realized profits. Primes continually struggle to provide sufficient justification for pricing, citing competition concerns or trade secrets, which in turn hinder our contracting professionals from assessing fair and reasonable pricing. To address this concern, the DoD should consider requiring prime contractors to formally notify contracting officers when they encounter difficulties in negotiations with their subcontractors. This notification should occur as soon as practicable to enable early intervention and facilitate alternative solutions into acquisition timelines. Increasing government awareness sooner rather than later, as currently seen, could help mitigate cost transparency and prevent government overpayment.

To formalize this requirement, the DoD could amend DFARS PGI 215.404-3, Subcontract Pricing Considerations, to mandate prime contractors to notify when cost data is being denied. Similarly, FAR 15.403-4, Requiring Cost or Pricing Data, could be amended to clarify that subcontractor refusal to provide cost data would be subject to additional reviews.

Alternative strategies could require contractors to justify cost reasonableness through independent audits or similar third-party verification. Formal enforcement by establishing a standardized template that prime contractors must submit to contracting officers detailing their efforts to obtain data as well as methodology for price reasonableness. To ensure compliance with either of these requirements, the DoD could implement penalties for non-compliance, such as withholding fees or reducing incentive payments in addition to making this requirement a key factor when completing Contractor Performance Assessment Reporting System (CPARS) ratings.

Expansion of Industrial Base Analysis & Sustainment Program

Industrial Base Analysis & Sustainment (IBAS) is an initiative that was established in 2014 under 10 U.S.C. § 2508 to (1) support the monitoring and assessment of the industrial base, (2) address critical issues in the industrial base relating to urgent operational needs, (3) support efforts to expand the industrial base, and (4) address supply chain vulnerabilities. Managed by the Office of the Deputy Assistant Secretary of Defense for Industrial Base Policy (ODASD[IBP]), the program plays a key role in identifying risks to our industrial base. In order to effectively address the impact of increasing horizontal integration within the defense industrial base, IBAS should systematically track deficiencies arising from subcontractor dependencies and the growing reliance on sole-source vendors. To fully leverage IBAS capabilities, it would be beneficial to assess the percentage (by both number and dollar value) of suppliers in the industrial base data that are sole-source, demonstrating the extent of reliance and emphasizing the potential impact of targeted interventions.

IBAS could map and track subcontractor relationships across major defense programs and pinpoint areas where M&As have limited competition. Concurrently, it could identify single points of failure where there have been significant supply chain disruptions. These disruptions could be due to production shortfalls, bankruptcies, or foreign acquisitions. Establishing or developing an early warning indicator framework could identify these reliances early on and help predict vulnerabilities. Tracking and analyzing these risks could help anticipate and mitigate these issues in future major defense spending. IBAS has historically been used to create new domestic sources, often resulting in sole-source suppliers where no domestic alternative existed. A significant and potentially transformative shift would be to strategically leverage IBAS to create and expand the pool of domestic vendors who can compete with each other, fostering a more resilient and cost-effective industrial base.

The program can fund initiatives which bolster the DoD industrial base. Currently it invests heavily in six priority industrial capability development areas. These are Submarine and Shipbuilding Workforce, Kinetic Weapons, Microelectronics, Critical Chemicals, Castings and Forgings, and Energy Storage and Batteries (OUSD[A&S], n.d.). Funding could be prioritized in finding alternative suppliers where consolidation has previously decreased competition, particularly in areas where sole-source reliance is demonstrably high. This funding could incentivize new entrants, increasing competition at the subcontractor level and mitigating the risks associated with concentrated supply chains.

Small Business Innovation Research Subcontractor Fast Track

The Small Business Innovation Research (SBIR) program was established in 1982 by Congress to stimulate technological innovation by providing research and development (R&D) funds to small businesses with 500 or fewer employees. A proposed Subcontractor Fast Track SBIR would target small businesses that could potentially become alternative suppliers for the prime defense industrial base. A SBIR of this nature could enable small businesses to scale their production in order to supply big defense contractors with critical components or services. Participants of this SBIR could potentially enhance competition, reduce the supply chain risk by diversifying subcontractor portfolios, and stimulate economic growth and advancements in the military defense sector.

In the interest of establishing a SBIR Subcontract Fast Track, it is important to first identify critical programs or components which already struggle with limited suppliers. After successful award to these small businesses, it is important to continue to provide support and resources beyond the contract. Offering technical assistance and mentorship is vital in ensuring they successfully develop and scale their ideas and innovations. Doing so would establish supplier diversity for critical defense needs and improved fresh innovative perspective to an already dated defense base.

Conclusion

Increased reliance of the DoD's prime contractors on subcontractors has heavily impacted and transformed the way the defense industrial base does business with the government. It has shifted cost structures and likely reduced transparency in government acquisitions. It is imperative that the government take action to ensure fair and reasonable pricing to protect taxpayer dollars. The consequences of unreformed consolidation demand policy intervention.

Our research has indicated that there has been an increase in material and subcontractor costs, in proportion to overall direct costs, over the past 15 years. Within the past 3 years, we found that material costs have averaged about 80%. Furthermore, our research data concluded that these costs for the same programs over time have also been on the rise. Since the consolidation of the defense industry over the past 30 years, there has been a shift in prime-contractor business models which has resulted in these prime-contractors subcontracting more work (in particular on the production of weapon systems) and concentrating on systems integration. These consolidations have led to an increasing reliance on a smaller number of contractors for critical defense capabilities. Consequently, promoting competition and ensuring it is fair and open for future programs should be a top priority.

One of the ways that contractors have taken control over their supply chain due to supply risk vulnerabilities, among other concerns, is to become a vertical integrator of their materials in two or more steps in the supply chain. When this process is done well, the benefits can include lower costs, greater control, and improved supply chain visibility. However, this can also lead to greater costs which comes from the upfront investment from acquiring or merging with suppliers, manufacturers, additional facilities and employees, and new business processes. Effectively, this also reduces competition, which helps to ensure that buying decisions are fair and objective. The future of the DoD will be shaped by steps taken now to increase competition and the number of suppliers in the defense industrial base.

In order to counter the rising shift in direct materials and subcontractor costs, we need to take necessary action, from promoting competition, expanding the defense industrial base, and reducing barriers for small businesses to compete to implementing policies that overall would increase our supplier diversity, reduce costs, and create innovative efficiencies. Imagine a future where our supply base thrives, fueled by healthy competition and a skilled workforce. To achieve this vision, we must delve deeper into understanding the dynamics of competition within our supplier network. Future research should specifically analyze competition rates within the supply base, examining the number of qualified suppliers vying for contracts at various tiers. Attracting and retaining top talent, while simultaneously bridging any skill gaps, will be crucial for fostering collaborative growth with all stakeholders. Controlling unit prices is another critical challenge. Building strong supplier relationships and negotiating advantageous terms are promising avenues to explore. Finally, dissecting unit prices based on acquisition type—sole source, competitive bids, off-the-shelf solutions, modified commercial products, and so on—will illuminate how different procurement strategies influence cost and unlock opportunities for optimization. This unit price analysis should further be broken down by acquisition type (sole source, competitive, COTS, etc.) to identify specific areas where cost control measures are most effective. This multifaceted approach will pave the way for a more robust and resilient supply base. Additionally, further investigation into contract types and their correlation with subcontracting costs is also warranted.

Disclaimer

The views expressed are those of the author(s) and do not reflect the official policy or position of the Naval Postgraduate School, US Navy, Department of Defense, or the US Government.

References

Congressional Research Service. (2023, December 19). *Department of Defense contract pricing* (R47879). https://crsreports.congress.gov

Defense Federal Acquisition Regulation Supplement 234.71, Cost and Software Data Reporting (2025). Acquisition.gov https://www.acquisition.gov/dfars/subpart-234.71-cost-and-software-data-reporting

FAR 42.202, Assignment of Contract Administration (2025). Acquisition.gov

GAO. (2022, July 7). *Defense industrial base: DoD should take actions to strengthen its risk mitigation approach*. https://www.gao.gov/products/gao-22-104154

Gowtham, S. (2024, January 8). *Cost control: Definition, techniques, methods, strategies & examples*. https://happay.com/blog/cost-control/

Greenwood Aerospace. (2023, June 23) *Regulating supply and demand: Prime government contractor's influence on aerospace procurement*. https://www.governmentprocurement.com/news/prime-government-contractors

Howard, T., et al. (2024, December 18). *NDAA provisions impacting government contractors and their supply chains*. https://www.wiley.law/alert-NDAA-Provisions-Impacting-Governments-Contractors-and-their-Supply-chains

Nicastro, L. (2024, September 23). *The US defense industrial base: Background and issues for Congress* (R47751). Congressional Research Service. https://www.congress.gov/crs-product/R47751

Office of the Under Secretary of Defense for Acquisition and Sustainment. (n.d.). *Assistant Secretary of Defense for Industrial Base Policy*. OUSD A&S - Industrial Base Policy. https://www.businessdefense.gov/ibr/mceip/icam/index.html

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

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics. (2016, March). *Guidance on using incentive and other contract types* (USA001270-16-DPAP). https://www.acq.osd.mil/dpap/policy/policyvault/USA001270-16-DPAP.pdf

Sanders, R. (2024, November 21) *Why supply chain cost reduction has become a boardroom discussion*. EY. https://www.ey.com/en_us?coo/why-supply-chain-cost-has-become-a-boardroom-discussion

Shinego, W. (2024, December 4). *DoD leverages Defense Production Act to galvanize critical supply chains*. DoD. https://www.defense.gov/news/news-stories/article/aricle/3985393/dod-leverages-defense-production-act-to-galvanize-critical-supply-chains/

Tirpak, J. A. (1998, July 1). The distillation of the defense industry. *Air & Space Forces Magazine*. https://www.airandspaceforces.com/article/0798industry/

Tirpak, J. A. (2022, February 16). New Pentagon report raises alarm over industry consolidation, future of competition. *Air & Space Forces Magazine*. https://www.airandspaceforces.com/new-pentagon-report-raises-alarm-over-industry-consolidation-future-of-competition/

Vergun, D. (2022, January 19). *DoD hopes for legislative action on product, services pricing policy*. DoD News. https://defense.gov/news/news-stories/article/article/2904198/dod-hopes-for-legislative-action-on-product-services-pricing-policy/

Supplier Chain Visibility—A Contracting Necessity

Rafaela Kovacs—is the Chief of Contracts at Air Force Operational Test and Evaluation Center, responsible for a contract portfolio providing relevant, credible, and timely test and evaluation to more than 90 major acquisition programs. Kovacs is also the Air Force Operational Test and Evaluation Center Director of Small Business Programs and responsible for managing the small business program on behalf of the commander. She holds an unlimited contracting officer's warrant, agreements officer under 10 USC 4022, and is a certified professional contracts manager. Kovacs earned her Master of Business Administration at the University of New Mexico. [Rafaela.a.kovacs.civ@mail.mil]

Tamie Haines—is a Supervisory Price and Cost Analyst with the Department of Contract Management Agency, having held various acquisition roles, including contract administrator and specialist, as well as a warranted Administrative Contracting Officer for prominent defense weapons systems. Haines served in the United States Air Force, commissioned through the Air National Guard in 2017, and is currently an A-Staff Officer. She previously served as a Space Control Officer and Contracting Officer with Flight Commander and Contingency experience. Haines holds a Master in Business Administration degree and is pursuing a doctorate in public administration. [Tamie.n.haines.civ@mail.mil]

Peter Guinto—Mentor

Abstract

This paper investigates the weaknesses present in federal procurement supply chains, emphasizing cost analysis, cybersecurity, and risk mitigation within defense contracts. Utilizing data from more than 200 Contractor Cost Data Reports and 87 Price Negotiation Memorandums (2015–2025), the analysis shows that subcontracting, materials, and inter-company transfers constitute more than 80% of direct costs, emphasizing the necessity for increased transparency and accountability in government acquisitions. The study points out significant threats, including dependence on foreign sources, cybersecurity vulnerabilities, and inadequate oversight of single-source suppliers. It assesses private-sector frameworks like IBM's cognitive supply chain and Starbucks' enterprise risk management to pinpoint best practices applicable to defense procurement.

This paper also considers unsuccessful legislative attempts to enforce supply chain compliance, highlighting the significance of flexible policies, contractor involvement, and financial incentives. Key recommendations include improving bill of materials transparency, widening the oversight of DCMA and DCAA, and bolstering domestic manufacturing efforts. Case studies from the F-35 and C-17 programs demonstrate the repercussions of insufficient oversight and the benefits of responsible sourcing practices. Ultimately, the paper promotes a forward-thinking, technology-focused, and risk-managed procurement approach that enhances national security and fortifies the resilience of the defense industrial base.

Introduction

Cost analysis in government contracting is essential for ensuring financial accountability, identifying inefficiencies, and optimizing expenditures. An analysis examining 212 Contractor Cost Data Reports (CCDRs) from 2015 through March 2025 revealed $974 million in subcontracting, material, and inter-company work transfers (IWT) expenses, with these categories comprising 83% of total direct costs (DoD, 2025a). Similarly, a review of 87 Defense Pricing and Contracting, Acquisition Policy (DPCAP) Price Negotiation Memorandums (PNMs) from 2022 to 2025 resulted in $120.7 billion in subcontracting, material, and IWT expenses, accounting for 81% of total direct costs (DoD, 2025b). The alignment between CCDR and PNM data highlights consistent cost distributions across programs, allowing a deeper examination of cost drivers and potential efficiencies within government acquisition.

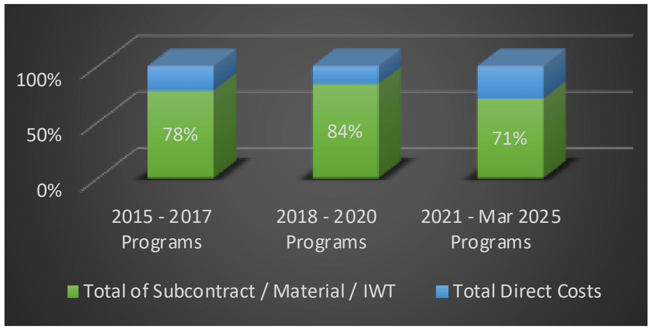


Figure 1. CCDRs 2022–2025

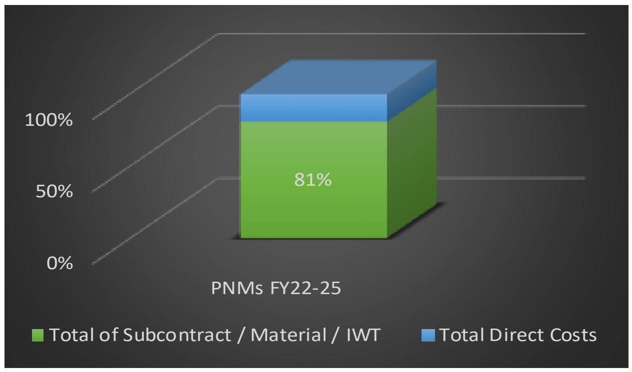


Figure 2. DPCAP PNMs Peer Reviews 2022–2025

Supply Chain Vulnerabilities and Cybersecurity Threats In Federal Procurement

The security of the federal supply chain is a critical concern in defense contracting, where reliance on foreign sources and cybersecurity threats create serious vulnerabilities. The Government Accountability Office (GAO) has identified weaknesses in managing risks, particularly concerning single-source suppliers and critical material dependencies. Rare earth elements essential for defense systems are largely sourced from foreign suppliers—especially China—raising concerns about supply disruptions and adversarial influence (GAO, 2024a).

The Department of Defense (DoD) also depends heavily on sole-source contractors for essential components. A GAO report found that the DoD does not consistently assess risks related to these suppliers, limiting its ability to prepare for disruptions (GAO, 2017). The report recommends more robust frameworks for tracking and mitigating supply chain dependencies.

Cybersecurity presents another major risk to federal procurement systems. Proposed amendments to the Federal Acquisition Regulation (FAR) would require vendors to report incidents within eight hours and disclose vulnerabilities in supply chain software (DoD et al., 2023). The Cybersecurity and Infrastructure Security Agency (CISA) further emphasizes the need for Software Bills of Materials (SBOMs) to identify and manage risks in third-party software (CISA, 2023).

Finally, the waiver process for domestic preference laws lacks consistent oversight and data accuracy. Although agencies are required to report when foreign goods are purchased, the GAO found errors and gaps in reporting that weaken transparency (GAO, 2024b). Improving this process would enhance accountability and help reinforce domestic sourcing in federal procurement.

Supply Chain Resilience and Disruption Mitigation

The Evolving Landscape Of Supply Chain Disruptions

The complexities of modern supply chains have increased vulnerabilities to disruptions, requiring strengthening mitigation strategies. Recent research identifies the severity of supply chain interruptions and examines frameworks organizations can adopt to enhance resilience. Blackhurst et al. (2005) recommend critical research areas in managing disruptions and identifying the need for proactive risk assessment and rapid recovery strategies. They contend that firms must develop a deeper understanding of supply chain vulnerabilities to effectively mitigate operational and financial risks. Additionally, Craighead et al. (2007) expand on this perspective, stating that the severity of supply chain disruptions is affected by design characteristics such as density, complexity, and node criticality. Their study emphasizes that supply chains with high node interdependence are more vulnerable to cascading failures, making it essential for companies to cultivate reactive and proactive mitigation capabilities.

Centers of Excellence as a Strategic Response

The idea of Supply Management Centers of Excellence (SM COEs) has surfaced as a systematic method for enhancing supply chain resilience (Handfield, 2024). A CAPS Research study outlines optimal practices for building COEs focused on supply market intelligence, risk management, and advanced analytics. Evidence indicates that organizations with dedicated COEs are more likely to standardize procurement procedures, improve category management insights, and employ data-driven decision-making models that foresee and address supply chain risks. The findings show that 59% of surveyed companies have at least one SM COE, with another 10% in the process of establishing one. These centers play a vital role in promoting best practices in supply management, disseminating intelligence, and encouraging a forward-thinking approach to risk management. Additionally, the research points to the growing incorporation of predictive analytics and digital twins in COEs to facilitate real-time monitoring and forecasting of supply chain disruptions.

Dynamic Stress Testing for Supply Chain Resilience

Stress testing is an emerging approach aimed at enhancing supply chain resilience, drawing parallels with established financial risk assessment techniques. Handfield et al. (n.d.) advocate for dynamic stress testing, which incorporates artificial intelligence (AI) and machine learning (ML) to deliver real-time scenario evaluations and predictive disruption notifications. Their research highlights the value of AI-powered simulations that consistently adjust risk assessments in response to changing global conditions. Through stress testing, organizations can proactively identify weaknesses by simulating possible supply chain shocks like geopolitical tensions, trade limitations, and resource scarcity. For instance, Honeywell has implemented dynamic stress testing to address risks associated with tariffs, international conflicts, and supply chain interruptions due to climate change. This methodology has enabled firms to bounce back more swiftly from disturbances, maintain supply continuity, and enhance supplier relationships through proactive oversight.

Build Resilience into the Procurement Process

Building a resilient supply chain requires a strategic approach focused on both downstream management and upstream vision. It is more than mere component acquisition. A procurement center can most effectively collaborate with risk management functions and other business units to manage supply chain risk (Schnellbacher et al., 2023). In one survey, only 10% of respondents indicated that their companies utilized a full range of capabilities to build a resilient supply chain.

Supply resilience can be achieved by combining human and technological resources. Three main activities can mitigate supply chain security threats from foreign dependencies and cyber vulnerabilities. Identifying hidden risks involves examining tier-2 suppliers for critical material sources, particularly rare earth elements, to reduce reliance on single foreign sources. Upon completing this analysis, AI systems can provide early warnings, recommend alternative sources, and strategize for strategic reserves while addressing vulnerabilities through strategic buffering, alternative supplier development, and stricter security standards. Enhancing cybersecurity during procurement through third-party security ratings or cybersecurity questionnaires enables proactive risk management and boosts the chain's overall resilience.

Lessons from the COVID-19 Supply Chain Crisis

The COVID-19 pandemic acted as a crucial examination of global supply chains, revealing critical gaps in visibility and responsiveness. Finkenstadt and Handfield (2021) investigate the visibility challenges faced in supply chains during the pandemic, particularly in sourcing personal protective equipment (PPE). Their findings show that dependence on low-cost suppliers from various regions exacerbated shortages and delays. The research highlights the need for supply chain mapping, diversified inventory, and investments in digital tracking technologies to improve visibility and readiness for future disruptions. The changing landscape of supply chain interruptions requires a strategic and layered approach to resilience. By integrating SM COEs, conducting dynamic stress tests, and employing AI-driven analytics, organizations can create an effective risk management framework. The lessons from the COVID-19 pandemic highlight the critical need for investment in tools for supply chain visibility and the establishment of strong mitigation strategies. As supply chains become more intricate, organizations must take proactive steps to implement these strategies, ensuring operational continuity and boosting their competitive edge.

F-35 Supply Chain Challenges and Security Risks

The F-35 Joint Strike Fighter program, the DoD's most expensive weapon system, has encountered ongoing supply chain issues, such as production delays, semiconductor shortages, and security risks linked to foreign-sourced materials (GAO, 2024c). One of the most concerning vulnerabilities was the discovery of a Chinese-sourced magnet within the aircraft's power system, raising alarms about adversarial infiltration and potential national security threats (Magnuson, 2022). The existence of foreign-manufactured components in critical defense systems highlights the necessity for tighter supply chain oversight and stronger sourcing policies to reduce security risks. A review of seven F-35 Lightning II and Air Vehicle Production acquisitions valuing $33.48 billion in total direct costs highlighted a significant 92% expenditure profile of subcontracting, material, and inter-company work transfers (IWT) expenses (DoD, 2025a, 2025b).

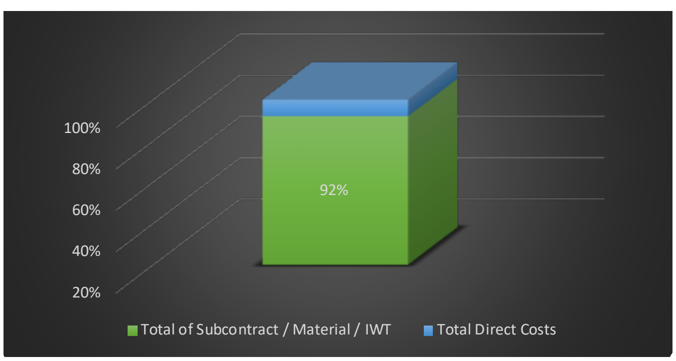


Figure 3. F-35 Expenditure Profile Review

Additionally, disruptions in the semiconductor supply chain have hindered F-35 production, worsening modernization delays and raising program costs (Fulco, 2023). The program's dependency on global semiconductor suppliers, mainly from geopolitically sensitive regions, has made it susceptible to supply shocks. Shivakumar and Wessner (2022) emphasize that semiconductors are crucial for national defense, and reliance on foreign sources poses operational risks that may affect military readiness.

A 2024 GAO report revealed that contractors deliver engines and aircraft late due to ongoing manufacturing issues and parts shortages. Technology Refresh 3 (TR-3), a $1.8 billion upgrade for the F-35's Block 4 modernization, faces delays from supply chain disruptions, including software and hardware shortages (GAO, 2024c). These delays impact cost efficiency and the DoD's ability to maintain a competitive technological edge. To address these challenges, the DoD must implement stronger supply chain visibility mechanisms, enforce stricter sourcing transparency requirements, and explore domestic semiconductor production to reduce foreign dependency. By adopting best practices from private sector supply chain management, such as blockchain tracking and AI-driven procurement monitoring, the DoD can improve oversight and mitigate F-35 supply chain risks.

C-17 Supply Chain Challenges

The C-17 Globemaster III, a critical aircraft for tactical airlift and airdrop missions and aeromedical evacuations, highlights the DoD's need to better manage supply chains for defense systems and platforms. A DoD Inspector General (IG) audit of its performance-based logistics (PBL) contracts exposed significant vulnerabilities in acquiring spare parts at fair and reasonable prices, stemming from the Department's handling of the bill of materials (BOM). A review of two C-17 Globemaster III acquisitions of $942 million in total direct costs highlighted a 57% expenditure profile of subcontracting, material, and IWT expenses.

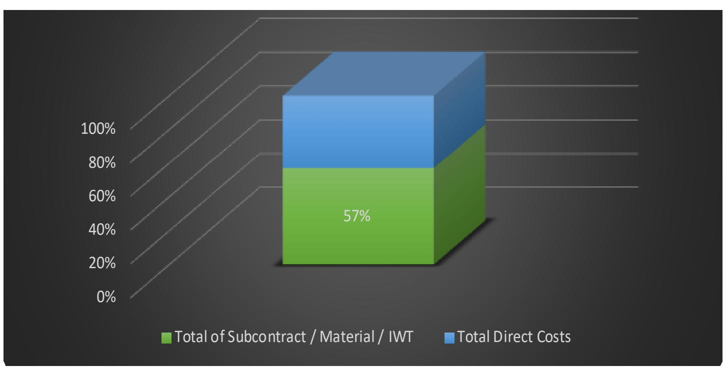


Figure 4. C-17 Globemaster III Expenditure Profile Review

Sole-source contracts like the C-17 PBL create an uneven playing field for negotiating prices. The vendor creates information asymmetry by the details it chooses to include in the BOM and cost data; thus, the government relies on the vendor's data to create its negotiation position. The lack of transparency and limited negotiation leverage increases the risk of inflated pricing.

Furthermore, the DoD does not require BOMs to be incorporated into the contract, which can create a disparity between proposed and actual materials. The audit found that 46.5% of the items delivered under the contract were included in the proposed BOM. Allowing vendors the discretion to provide materials of their choice undermines the initial determination of fair and price reasonableness, makes it difficult to anticipate and mitigate risks associated with diminishing manufacturing sources and material shortages (DoD, Office of Inspector General, 2024), and introduces potential quality control issues that affect readiness and safety.

Two approaches can be utilized in efforts to resolve audit findings. First, clear requirements for submitting complete and accurate BOMs would enhance total supply chain visibility. Second, strategies to reduce reliance on sole-source contracts could strengthen the negotiation position, such as incentivizing competition through dual sourcing or an open bidding process or proactively developing alternative sources for critical components or materials. Ultimately, improved data transparency would provide the Government with visibility into vendor pricing data and subcontractor relationships.

Lastly, retaining the Design Control Authority (DCA) is another targeted approach that could be applied in specific situations to mitigate challenges. The DCA is most appropriate for aircraft or weapons system programs where changes significantly affect components, manufacturing, and overall supply chain stability. It assists in managing obsolescence and mitigating supply chain risks that threaten the mission or national security. Regardless of the mechanism, addressing these supply chain vulnerabilities will position the DoDto ensure that the C-17 and other critical assets remain mission-ready while maintaining responsible stewardship of taxpayer resources.

Raytheon Settlement: A Cautionary Case for Subcontractor Oversight

The recent $950 million settlement between Raytheon Company and the U.S. Department of Justice (DOJ) highlights the risks of managing subcontractors and suppliers in federal procurement. The DOJ reported that Raytheon's subsidiary, RTX Corporation (previously known as Raytheon Technologies), admitted to participating in a bribery and fraud scheme that lasted a decade, which involved its jet engine manufacturer, Pratt & Whitney. This scheme included the establishment of fake subcontracts that funneled more than $55 million in bribes to government officials across various foreign nations to obtain defense contracts. Additionally, there were instances where Raytheon employees submitted false or misleading certifications, leading to the export of sensitive military hardware and technology of U.S. origin to unauthorized entities, thus breaching the Arms Export Control Act (AECA) and the International Traffic in Arms Regulations (ITAR; U.S. Department of Justice, 2024).

Using false subcontracts to disguise bribery payments highlights significant internal controls, subcontractor evaluation, and export compliance failures. Additionally, it raises serious doubts about supplier oversight within the defense industrial sector, particularly concerning high-risk components such as jet engines and aerospace technologies. In instances like the C-17 program, where sole-source vendors prevail, and prime contractors maintain information imbalances, this situation emphasizes strong oversight systems that authenticate subcontractor credibility, promote transparent billing practices, and reduce the risk of corruption and export violations. Enhancing these controls during the acquisition phase, instead of relying exclusively on subsequent sustainment audits, would allow for earlier identification and prevention of such misconduct.

Failed Legislative Attempts to Incentivize Supply Chain Compliance

Several legislative efforts to enhance transparency in defense contracting supply chains have faced obstacles linked to feasibility, cost, and enforcement. For example, the Supply Chain Illumination provision was initially framed for defense contractors to implement supply chain monitoring technologies and gave a short period for the Secretary of Defense to create incentives and minimum technical standards, including cybersecurity requirements. Objections to specific tool mandates on the grounds of possible security threats and small business compliance strains resulted in a requirement to incentive contractors to assess and monitor the entire supply chain for potential vulnerabilities and noncompliance risks (H. R. 5009, § 849). Another initiative to reimburse contractors for unforeseen disruptions and introduced a quick waiver process stalled due to budgetary issues and worries that contractors might pass costs onto the government. Additionally, a proposal to expedite supply chain reviews, referencing 10 U.S.C. § 4863 and 4872, aimed to improve risk disclosures concerning specialty metals and restricted foreign materials by providing a temporary National Security Waiver (NSW) as a corrective measure. Contractor hesitancy to self-report non-compliance due to risk of penalties, and conflicts with existing procurement regulations halted proposal consideration. Lastly, a proposal for safe harbor encouraged contractor transparency (report supply chain weakness and non-compliant materials) by protecting the disclosures from penalties and accepting non-compliant materials during NSW reviews. Critics contended this would weaken enforcement by shifting liability to the government and diminishing contractor accountability. These unsuccessful legislative attempts highlight the complexity needed to balance effective and enforceable supply chain reform with compliance requirements, cost implications, and contractor responsibility.

Compliance requirements such as mandating specific tools encountered substantial industry pushback, especially from small enterprises. Objections to the initial Supply Chain Illumination draft voiced vendor dependency and that inflexible standards and tight deadlines offered limited flexibility for scalable implementation. There would likely be inconsistent implementation across the defense industrial sector. Waiver-based compliance model initiatives, like expedited supply chain review, frequently clash with procurement regulations, leading to procedural delays and reduced contractor involvement. Further, the potential repercussions of voluntary disclosures created hesitancy towards transparency, undermining the purpose of the interim waiver system, which aimed to promote corrective actions and accountability.

Cost considerations hampered several of proposed legislation efforts. Hefty costs associated with compliance was another protest to the first draft of Supply Chain Illumination. The Department's budget constraints hindered ability to provide financial incentives intended to promote stronger supply chain management. Uncertainty about cost responsibilities discourages contractors from committing to strong risk management strategies.

Contractor responsibility is paramount in supply chain management. Although liability protections like a safe harbor encourage transparency, they often appear to reduce contractor accountability. From a broader industrial viewpoint, the decline of domestic capabilities, particularly in rare earth magnet manufacturing, poses challenges for compliance, sometimes placing it beyond a contractor's immediate influence. Consequently, many companies are reluctant to divulge proprietary information to potential rivals or "competimates," worried about losing their competitive edge or facing disintermediation. Additionally, cash flow is a vital consideration in these decisions, as disruptions in network flow significantly impact compliance and performance results. Ultimately, it is crucial for the government to respond promptly, not just to control costs and ensure taxpayer responsibility but also to foster the operational stability contractors require to sustain profitability.

Restoring Freedom's Forge Act: A Path Toward Supply Chain Modernization

The Restoring Freedom's Forge Act represents a legislative effort to revitalize defense procurement and enhance supply chain resilience (Restoring Freedom's Forge Act, 2024) by streamlining procurement processes, eliminating bureaucratic barriers, and strengthening domestic manufacturing capabilities. It attempts to resolve DoD procurement challenges like inadequate supply chain transparency and excessive dependence on foreign materials through mandates for better tracking and reporting methods to improve supply chain transparency.

Issues highlighted in recent GAO reports (2024a, 2024b, 2024c) include single-source reliance, cybersecurity risks, and uneven enforcement of domestic sourcing regulations. To tackle these shortcomings, the legislation outlines a three-part strategy. First, it aims to eliminate bureaucratic hurdles that hinder procurement and limit competition, thus speeding up the acquisition process and expanding the supplier pool (Restoring Freedom's Forge Act, 2024). Second, it intends to boost supply chain visibility by requiring enhanced tracking and reporting systems, enabling government agencies to pinpoint vulnerabilities sooner and respond more accurately. Finally, the act encourages domestic production by providing incentives for the onshoring of essential materials and components, thereby reducing U.S. dependence on potentially hostile nations for crucial defense supplies. Together, these actions represent a concerted effort to modernize federal procurement and strengthen national security through more robust and accountable supply chains.

One of the act's key provisions focuses on streamlining acquisition regulations, aligning with previous reports' recommendations highlighting the adverse effects of excessive oversight and slow procurement cycles (Restoring Freedom's Forge Act, 2024). By simplifying the approval process for new defense suppliers, the legislation aims to diversify the DoD's supply base, reducing the risks associated with single-source suppliers (GAO, 2017). Additionally, the act encourages companies to produce critical materials onshore, especially in sectors such as rare earth elements and semiconductors, which are essential for national security (GAO, 2024a).

Another important aspect of the act is incorporating modern technology into procurement oversight. By utilizing blockchain, AI-driven supply chain monitoring, and digital ledger tracking, the act seeks to enhance supply chain transparency (Restoring Freedom's Forge Act, 2024). These tools can help reduce cybersecurity threats and improve real-time tracking of subcontractor performance, strengthening recent amendments to the FAR (DoD et al., 2023).

The Act is likely to face similar challenges as the failed legislative attempts especially with contractors' full compliance with enhanced supply chain reporting requirements. There is industry resistance to increased regulatory burdens. Additionally, critics contend that excessive procurement deregulation could lessen accountability, and raise the risks of fraud, cost overruns, and security vulnerabilities. While the act promotes domestic manufacturing, enhancing U.S. production capacity for critical materials will demand time and significant investment (Shivakumar & Wessner, 2022). To ensure effectiveness, streamlining procurement and maintaining adequate oversight must be balanced with adequate measures to prevent supply chain vulnerabilities.

The Restoring Freedom's Forge Act represents a forward-thinking approach to defense procurement reform, tackling supply chain inefficiencies and bolstering domestic manufacturing (Restoring Freedom's Forge Act, 2024).While the act can potentially improve the DoD's supply chain resilience, its success will hinge on practical implementation, industry compliance, and consistent investment in domestic production capabilities. Future research should analyze how effectively the act meets its objectives and whether additional safeguards are required to mitigate potential risks in a less regulated procurement environment.

Private Industries' Success in Supply Chain Management

IBM's Supply Chain Management: A Model For Government Adoption

IBM has established itself as a leader in supply chain management by leveraging AI, predictive analytics, and sustainability-driven strategies. By implementing a cognitive supply chain, IBM has achieved 100% order fulfillment and cost savings of $160 million, demonstrating the effectiveness of AI-powered decision-making and risk mitigation (Martinez, 2023). The federal government, particularly the DoD and procurement agencies, could adopt IBM's methodologies to enhance supply chain visibility, resilience, and efficiency.

IBM's cognitive supply chain incorporates AI, ML, and real-time data analytics to optimize logistics and procurement. This system enables IBM to proactively address disruptions by predicting demand, optimizing inventory, and mitigating supplier risks (Martinez, 2023). The federal government, facing challenges with supply chain transparency, could benefit from a similar AI-based approach to tracking spending, identifying supply vulnerabilities, and improving contract management.

IBM incorporates sustainability into its supply chain by utilizing responsible sourcing and conducting environmental impact assessments. This structure ensures compliance with international regulations while enhancing operational efficiency (McGrath & Jonker, 2024). The DoD and other agencies responsible for managing critical materials could adopt IBM's sustainable supply chain model to improve resilience and security, especially in overseeing semiconductor and rare-earth element supplies.

One of IBM's key innovations is its ability to consolidate diverse legacy systems into a unified, transparent supply chain platform (IBM, n.d.). This cohesive approach allows all stakeholders to access the same real-time data, enhancing coordination between suppliers, logistics providers, and procurement teams. The government could adopt a similar digital infrastructure to increase visibility at the subcontractor level, mitigating risks associated with unverified foreign suppliers.

Starbucks Supply Chain Management: Valuable Lessons for Government

Starbucks stands out in supply chain management, evolving with the global environment. From a single store in Seattle in 1971, it now boasts 40,000 locations. Over the past 17 years, Starbucks has reduced its footprint, cut logistics costs, improved logistics quality, and embraced new technology (Tabansi, 2023). It implemented enterprise resource management (ERM) to analyze macro trends related to materials, geopolitical events, and environmental changes, allowing for effective risk mitigation. These efforts provide the DoD insights to enhance transparency, expand the supplier base, and integrate new technologies.

One of Starbucks' initial supply chain efforts was investigating the cause of rising costs. The company discovered that outsourcing decisions had not been reassessed during the growth period, leading to an unnecessarily complex supply chain and a critical reorganization needed (O'Byrne, 2020). The DoD could adopt a similar strategy by requiring contractors to provide a detailed supply chain map, including all sub-tiers, to understand better the risks hidden within it.

Starbucks' supply chain management has incorporated updated technology as new capabilities have become available. In 2014, the company began utilizing enterprise resource management to identify macro trends that could potentially disrupt supply chain operations (Supply Chain Quarterly, 2014). Post-COVID, Starbucks integrated a new system that actively tracks risks in the supply chain and mitigates them to the extent that is within the company's control (SFK Inc. et al., 2024). The DoD could replicate this approach by investing in a platform that centralizes supply chain data across all its programs to aggregate, monitor, and analyze risks using predictive analytics.

Part of Starbucks' supply chain reorganization involved terminating ineffective partnerships and requiring weekly scorecards on service quality. Over two years, these efforts saved the company more than $500 million (O'Byrne, 2020). The DoD cannot dictate which prime or sub-tier contractors to include in its supply chain. However, it could adopt a similar approach by collaborating with the industry to develop standard quality levels that can be tracked in the aforementioned supply chain platform. This would be a public-private partnership to standardize data, necessitating cybersecurity measures to protect the supply chain information. The DoD could use the data to incentivize its prime contractors to reward sub-tier contractors that exceed quality levels or proactively manage risks down the supply chain.

AI in the Modern Supply Chain

AI is transforming all business sectors with its ability to analyze large data sets and identify complex patterns, creating the potential for enhanced decision-making, process optimization, and mitigating risks in an intricate global supply chain. The move towards AI supply chain management appears to be a strong operational decision backed by financial outcomes. In 2022, companies reported that costs decreased by more than 10% and revenue increased by more than 5% after their first year of AI adoption for supply chain management (Chui et al., 2022).

Benefits

Georgetown University's Walsh School of Foreign Service investigated AI's ability to develop a resilient supply chain and concluded its use has three advantages (Cohen & Tang, 2024). The capability to process vast amounts of data can predict fluctuations in demand with higher accuracy than historical methods, resulting in rightsized inventory levels. Data integration from diverse sources (supplier databases, news feeds, and social media analysis) builds a comprehensive view of the supply chain. The enhanced visibility allows for early identification of potential disruptions (supplier-related issues, geopolitical instability, or natural disasters). It equips decision-makers with time to develop and implement mitigation strategies to minimize chain disruptions and facilitates the evaluation of multiple scenario responses through simulations. By modeling the impact of different decisions, identifying the most effective solution becomes more transparent.

Limitations

While AI offers significant advantages in supply chain management, it is crucial to recognize its limitations. The efficacy of AI depends on the quality and accuracy of data inputs; as noted by an academic expert with extensive experience in Federal procurements and acquisitions, commercially available supply chain analytics utilize AI and ML to analyze publicly accessible information, which can result in inaccurate predictions and poor decision-making. These adverse outcomes may stem from the platform's inability to differentiate between outsourced and subcontracted relationships or between headquarters and plant locations. These limitations underscore the necessity of human oversight, especially in complex supply chain relationships (R. Handfield, personal communication, March 5, 2025).

Proposed Solutions to Enhance Supply Chain Oversight

A comprehensive set of policy, technological, and structural reforms must be adopted to address the persistent challenges in defense procurement and supply chain vulnerabilities. In addition to the lessons learned noted in previous sections, the recommendations aim to enhance transparency, mitigate risks, and strengthen domestic manufacturing capabilities, ensuring that national security interests are prioritized in the supply chain.

Recommendation 1: Strengthening Policy and Regulatory Frameworks

A major issue in defense procurement is the lack of visibility into lower-tier suppliers, heightening risks of foreign infiltration and counterfeit materials (GAO, 2024a). Revisions to the FAR and DFARS should ensure full transparency of lower-tier suppliers, especially for contracts involving critical components. Expanding DFARS 252.244-7001 to mandate disclosures from Tier 2+ suppliers would compel prime contractors to report subcontractor sources, preventing reliance on unverified foreign entities (Restoring Freedom's Forge Act, 2024). Annual supply chain reports from prime contractors should also be required to assess compliance and procurement integrity.

Another key policy reform seeks to limit exceptions for foreign sourcing by tightening waivers under the Berry Amendment and establishing domestic sourcing requirements for defense materials. The Defense Production Act (DPA) Title III should be managed proactively rather than reactively, with strategic long-term investments informed by forward-looking data mapping. While DPA Title III offers tools such as loan guarantees, direct purchases, and grants to expand domestic capacity (Office of the Assistant Secretary of Defense for Industrial Base Policy, n.d.), its full potential remains underutilized when implemented solely in response to crises.

For instance, China's 2023 export restrictions on gallium and germanium, a decision that sparked concern across the defense, energy, and electronics sectors, demonstrate how supply disruptions can emerge suddenly and at scale, affecting critical defense programs (Holderness et al., 2023). Increased use of predictive analytics and industrial base mapping could aid in identifying and addressing supply chain vulnerabilities sooner, ensuring funding is directed to stabilize domestic production before strategic materials become unavailable. These strategies would help lessen reliance on foreign-made materials while bolstering domestic industrial capacity and resilience.

Recommendation 2: Expanding Oversight and Workforce Training

Expanding the authority of the Defense Contract Management Agency (DCMA) and the Defense Contract Audit Agency (DCAA) for auditing will enhance oversight of subcontractors and ensure compliance with supply chain security policies. The 2024 GAO report highlights that inadequate auditing has led to supply chain inefficiencies and security vulnerabilities (GAO, 2024c). Strengthening DCMA's role in contract execution will provide greater enforcement capabilities to ensure that suppliers adhere to domestic sourcing and cybersecurity standards. Additionally, acquisition professionals need specialized training in supply chain risk management. Mandating training in supply chain security for contracting officers and program managers will enhance their ability to assess contractor compliance and mitigate risks related to foreign dependencies (Restoring Freedom's Forge Act, 2024).

Building on this, DCMA should also act as a central arbitrator to manage shared supplier resources across the defense industrial base. Without centralized coordination, prime contractors function like independent herders in the "tragedy of the commons" scenario, exhausting shared supplier capacity without insight into each other's activities (Broga, 2006; Investopedia, 2023). This absence of communication results in overbooking of suppliers, delayed deliveries, and inflated costs, burdens that ultimately fall on the government to absorb. The government assumes full system risk when subcontractors are overextended across multiple primes, and their limitations go unrecognized. To prevent this, supply chain oversight should take place within acquisition, rather than sustainment, to influence contract decisions before a crisis point. In this model, DCMA would enforce compliance and manage capacity transparency, ensuring sustainable use of industrial resources for national defense.

Recommendation 3: Enhance Bom Transparency and Responsible Sourcing Oversight

One of the most overlooked yet critical tools in supply chain risk management is the BOM. According to a Fortune 500 Chief Procurement Officer (CPO), BOMs are foundational to world-class supply chain management, and even executive-level leaders regularly review them due to their strategic importance (CPO, personal communication, March 7, 2025). Oversight of BOMs enables early identification of sourcing risks, particularly when the government has visibility into all levels of sub-tier suppliers, not just direct contractors.

The government must ensure that BOM reviews include a comprehensive understanding of the original sources of parts, particularly for critical components. Integrating emerging technologies into BOM and inventory analysis would improve visibility, integrity, and real-time tracking throughout the supply chain (CPO, personal communication, March 7, 2025). These tools can help prevent counterfeit parts, identify foreign vulnerabilities, and support proactive rather than reactive supply chain decisions.

The mutual dependency between government agencies and suppliers requires a collaborative and secure oversight model. Experienced supply chain subject matter experts (SMEs) could be engaged under non-disclosure agreements (NDAs) to guide BOM assessments, engineering change management, and overall procurement strategy to enable effective governance while maintaining confidentiality. This approach would protect proprietary information while ensuring expert insights inform acquisition decisions (CPO, personal communication, March 7, 2025).

Additionally, utilizing impartial third-party organizations, such as those following the Electronic Industry Citizenship Coalition (EICC) structure and the Responsible Business Alliance (RBA), can support ethical sourcing and help establish fair and reasonable pricing in contract negotiations. These entities already provide responsible sourcing verification and pricing analytics to the private sector, and their neutrality could enhance credibility and consistency in federal acquisition processes (CPO, personal communication, March 7, 2025). Integrating their capabilities would align defense procurement with commercial best practices while reinforcing transparency and sustainability across the defense industrial base.

Conclusion

An in-depth analysis of contractor cost data and procurement records reveals a pressing concern: the federal government lacks the necessary visibility into the subcontracting and material flows that comprise the backbone of our national defense supply chain. This systemic blind spot undermines strategic oversight, impedes proactive risk management, and jeopardizes fiscal responsibility and mission readiness. With subcontracting, material costs, and transfers between companies accounting for more than 80% of total direct costs, the lack of transparent oversight threatens taxpayer money and mission preparedness. Additionally, ongoing vulnerabilities discussed further jeopardize the integrity and resilience of the supply chain.

Insights gained from legislative missteps, successful private sector examples from companies like IBM and Starbucks, and notable defense projects, including the F-35 and C-17, indicate that the government's predominant emphasis on sustainment is inadequate. It is crucial to integrate risk management, data analytics, and supplier accountability much earlier in the life cycle to enhance procurement processes. For effective modernization and security of federal supply chains, the government must shift its focus to visibility during the acquisition stage, utilizing methods such as dynamic stress testing, predictive analytics, and AI-driven mapping to identify and mitigate threats proactively.

Moving forward requires more than just temporary solutions. It calls for a fundamental change in culture and operations regarding acquisition, policy, and oversight. This shift must focus on real-time transparency, enforce ethical sourcing standards, and encourage proactive teamwork within the defense industrial sector. Only then will the government be able to guarantee sturdy, efficient, and secure supply chains that address the changing needs of national security and fiscal accountability.

Disclaimer

The views expressed are those of the author(s) and do not reflect the official policy or position of the Naval Postgraduate School, US Navy, Department of Defense, or the US Government.

List of Acronyms and Abbreviations

3PL	Third-Party Logistics
AECA	Arms Export Control Act (AECA)
AI	Artificial Intelligence
BOM	Bill of Material
CCDR	Contractor Cost Data Report
CISA	Cybersecurity and Infrastructure Security Agency
CMMC	Cybersecurity Maturity Model Certification
COE	Centers of Excellence
CPO	Chief Product Officer
DCAA	Defense Contract Accounting Agency
DCMA	Defense Contract Management Agency
DoD	Department of Defense
DOJ	Department of Justice
DPA	Defense Production Act
EICC	Electronic Industry Citizenship Coalition
ERM	Enterprise Resource Management
FAR	Federal Acquisition Regulations
GAO	Government Accountability Office
DCA	Design Control Authority
DFAR	Defense Federal Acquisition Regulations
DPCAP	Defense Pricing Contracts and Acquisition Policy
IG	Inspector General
IoT	Internet of Things
ITAR	International Traffic in Arms Regulations
IWT	Inter-Work Transfers
ML	Machine Learning
NDA	Non-Disclosure Statement
NSW	National Security Waiver
PBL	Performance-Based Logistics
PNM	Price Negotiation Memorandum
PPE	Personal Protective Equipment
RBA	Responsible Business Alliance
RFID	Radio-Frequency Identification
SBOM	Software Bill of Materials
SM	Supply Management
SME	Subject Matter Expert
TR-3	Technology Refresh 3

List of References

Blackhurst, J., Craighead, C. W., Elkins, D., & Handfield, R. B. (2005). An empirically derived agenda of critical research issues for managing supply-chain disruptions. *International Journal of Production Research, 43*(19), 4067–4081. https://doi.org/10.1080/00207540500151549

Broga, C. (2006, Fall). *Jargon alert: Tragedy of the commons*. Region Focus. Federal Reserve Bank of Richmond. https://www.richmondfed.org/-/media/richmondfedorg/publications/research/econ_focus/2006/fall/pdf/jargon_alert.pdf

Chui, M., Hall, B., Mayhew, H., Singla, A., & Sukharevsky, A. (2022). *The state of AI in 2022—and a half-decade in review.* McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review

Cohen, M. C., & Tang, C. S. (2024, February 5). The role of AI in developing resilient supply chains. *Georgetown Journal of International Affairs.* https://gjia.georgetown.edu/2024/02/05/the-role-of-ai-in-developing-resilient-supply-chains/

Cybersecurity and Infrastructure Security Agency. (2023). *2023 top routinely exploited vulnerabilities.* https://www.cisa.gov/news-events/cybersecurity-advisories/aa24-317a

Craighead, C. W., Blackhurst, J., Rungtusanatham, M. J., & Handfield, R. B. (2007, February). The severity of supply chain disruptions: Design characteristics and mitigation capabilities. *Decision Sciences, 38*(1), 131–156. https://doi.org/10.1111/j.1540-5915.2007.00151.x

Finkenstadt, D. J., & Handfield, R. (2021). Blurry vision: Supply chain visibility for personal protective equipment during COVID-19. *Journal of Purchasing and Supply Management, 27*(3), 100689. https://doi.org/10.1016/j.pursup.2021.100689

DoD. (2025a). *Contractor Acquired Data Entry (CADE) system dataset* [Unpublished internal database].

DoD. (2025b). *Defense Contracting Pricing and Acquisition Policy, Price Negotiation Memorandums dataset* [Unpublished internal database].

DoD, GSA, & NASA. (2023). *Federal acquisition regulation: Cyber threat and incident reporting and information sharing.* https://www.federalregister.gov/documents/2023/10/03/2023-21328/federal-acquisition-regulation-cyber-threat-and-incident-reporting-and-information-sharing

Department of Defense, Office of Inspector General. (2024). *Audit of C-17 spare parts pricing* (Report No. DODIG-2025-009) [Audit report]. Department of Defense.

Fulco, M. (2023, June 13). How aerospace can improve its supply of semiconductors. *Aviation Week Network.* https://aviationweek.com/shownews/paris-air-show/how-aerospace-can-improve-its-supply-semiconductors

GAO. (2017). *Defense supply chain: DOD needs complete information on single sources of supply to proactively manage the risks* (GAO-17-768). https://www.gao.gov/products/GAO-17-768

GAO. (2024a). *Critical materials: Action needed to implement requirements that reduce supply chain risks* (GAO-24-107176). https://www.gao.gov/products/GAO-24-107176

GAO. (2024b). *Federal contracting: Opportunities exist to improve the reporting of waivers to domestic preference laws* (GAO-24-106166). https://www.gao.gov/products/GAO-24-106166

GAO. (2024c). *F-35 Joint strike fighter: Program continues to encounter production issues and modernization delays.* https://www.gao.gov/products/gao-24-106909

Handfield, R. (2024, January 23). *Best practices in forming centers of excellence*. CAPS Research. https://www.capsresearch.org/blog/posts/2024/january/caps-research-best-practices-in-forming-centers-of-excellence/

Handfield, R., Retherford, B., & DeGrange, W. (n.d.). *Why dynamic stress testing is critical for supply chain resilience.* NC State Poole College of Management. [Manuscript in preparation].

Holderness, A., Velazquez, N., Carroll, H. H., & Cook, C. (2023, July 7). *Understanding China's gallium sanctions*. Center for Strategic & International Studies. https://www.csis.org/analysis/understanding-chinas-gallium-sanctions

IBM. (n.d.). *IBM supply chain solutions.* https://www.ibm.com/supply-chain

Investopedia. (2023, September 25). *What is the tragedy of the commons in economics?* https://www.investopedia.com/terms/t/tragedy-of-the-commons.asp

Magnuson, S. (2022, September 20). AFA news: Chinese-sourced magnet in F-35 prompts supply chain concerns. *National Defense.* https://www.nationaldefensemagazine.org/articles/2022/9/20/chinese-sourced-magnet-in-f-35-prompts-supply-chain-concerns

Martinez, J. (2023, October 25). *IBM saves $160 million, achieves 100% order fulfillment with cognitive supply chain*. Cloud Wars. https://cloudwars.com/cloud/how-ibm-achieved-100-order-fulfillment-and-160-million-cost-savings-with-cognitive-supply-chain/

McGrath, A., & Jonker, A. (2024, March 14). *What is sustainable supply chain management?* IBM. https://www.ibm.com/think/topics/sustainable-supply-chain-management

O'Byrne, R. (2020, February 4). *7 mini case studies: Successful supply chain cost reduction and management*. https://www.linkedin.com/pulse/7-mini-case-studies-successful-supply-chain-cost-rob-o-byrne

Office of the Assistant Secretary of Defense for Industrial Base Policy. (n.d.). *Defense production act title III overview*. Retrieved from https://www.businessdefense.gov/ibr/mceip/dpai/dpat3/docs/DPA-TitleIII-Overview.pdf

SFK Inc., SKK Marine, & SFK SecCon. (2024, July 19). *Starbucks supply chain management: Optimizing global coffee distribution through risk mitigation and sustainable practices.* https://sfkcorp.com/starbucks-supply-chain-management-optimizing-global-coffee-distribution-through-risk-mitigation-and-sustainable-practices/

Shivakumar, S., & Wessner, C. (2022, June 8). *Semiconductors and national defense: What are the stakes?* Center for Strategic & International Studies. https://www.csis.org/analysis/semiconductors-and-national-defense-what-are-stakes

Supply Chain Quarterly. (2014, February 25). Starbucks adds risk management program to help protect its supply chain. https://www.thescxchange.com/finance-strategy/plan/risk-management-starbucks

Tabansi, O. (2023, December 1). *Starbucks's supply chain challenges and how it overcame them*. https://supplychainnuggets.com/starbuckss-supply-chain-challenges-and-how-it-overcame-them/

U.S. Department of Justice. (2024, October 16). *Raytheon company to pay over $950 million in connection with foreign bribery, export control, and fraud schemes*. https://www.justice.gov/usao-edny/pr/raytheon-company-pay-over-950-million-connection-foreign-bribery-export-control-and

U.S. Senate. (2024). *Restoring freedom's forge act.* Wicker Senate Office. https://www.wicker.senate.gov/services/files/4396C3A9-DA26-4BD6-A655-9E0910B83DA8

Commerciality: Real Savings?

Robert Hazlewood—is a Senior Quality Auditor in the Policy, Quality, and Legislative Affairs (PQLA) Directorate at the Defense Contract Audit Agency. Hazlewood leads teams of quality auditors (QAs) performing reviews of audits for compliance with Generally Accepted Government Auditing Standards (GAGAS), AICPA attestation standards (AT-Cs), and Agency policy. He assists in developing the annual quality review plan, conducting quality reviews, and works with DCAA Policy staff on matters impacting audit quality.

In 2024, he was selected to assist with a strategic improvement group (SIG) to address changes needed to comply with the 2024 Yellow Book. Hazlewood provided input throughout the process, covering every aspect of DCAA's system for ensuring audit quality. He was integral to mapping existing and "to-be" post-issuance review processes for audit engagements.

Prior to serving in PQLA Hazlewood was a Branch Manager in DCAA, overseeing the operations of field audit offices from 2012 to 2020. He was a Supervisory Auditor from 2003 to 2012, including a five-year tour in the European Branch Office in Wiesbaden, Germany. Hazlewood has also served as a Financial Liaison Advisor at Tinker AFB, OK from 1999 to 2003.

Hazlewood holds a Bachelor of Accountancy and Master of Public Administration Degrees from the University of Oklahoma. He completed the Director's Development Program in Leadership in 2019, and holds Level III certifications in DAWIA, FM, and DoD Defense Acquisition Corps. Hazlewood is a CPA in the State of Oklahoma since January 1990. [robert.c.hazlewood.civ@mail.mil]

Cher Mount— serves as a Contract Cost/Price Analyst within Team ACBB of the Cost and Pricing Command of DCMA. Her responsibilities include managing and reviewing a portfolio of proposals, issuing recommendations to proposals, and conducts rate and progress payment reviews for the Armed Forces Contracting Offices and teams within DCMA. In addition, Mount held a temporary Supervisor Position for Team ACBB in 2023. Her responsibilities included, managing the workload for the team to ensure proper assignment, review, and oversight. Prior to working within the Cost and Pricing Command, Mount has experience working as a Contractor Administrator with DCMA. Her responsibilities inncluded contract administration services such as assisting in the negotiating of un definitized contract actions, review, and closeout.

Prior to her responsibilities as a Contract Cost/Price Analyst, Mount began her career working as a Contracting Officer for Army NAF Contracting in Fort Campbell, Kentucky. She was responsible for the cradle to grave of contracting administration actions to include construction projects within the Southeast Region.

Mount has earned a Bachelor of Science in Business Administration and a Bachelor of Arts in Elementary Education from Point Park University in Pittsburgh, Pennsylvania. In 2017, Mount earned her MBA with a concentration in Public Administration from Columbia Southern University in Orange Beach, Alabama.

In her personal capacity and as an advocate for veterans, she volunteers for the nonprofit, Tails of Valor/Paws of Honor in Coopersburg, Pennsylvania as the Secretary of the Board of Directors. [cher.e.mount.civ@mail.mil]

Abstract

Commercial item pricing is viewed as a way to reduce costs to the Government. This has not been meaningfully assessed in relation to "of a type" commercial items where there is not a competitive commercial marketplace to shape prices. This research and panel would assess whether this type of part pricing is creating savings.

Introduction

Commercial, Commercial, Commercial. "Law has directed a preference for commercial item procurement since the early 1990s" (DoD, 2019, p. 4). This paper outlines a specific precept behind this preference (i.e., money savings) and examines whether the expected benefit has manifested.

Do costs decrease when products are treated as commercial? Proponents of classifying items as commercial point to fiscal savings promised by not having to follow FAR Part 15 and Cost Accounting Standards. For example, the regulations provide for obtaining cost data only if needed to determine a fair and reasonable price. The information provided by the offeror on sales to other commercial customers, market analysis by the acquisition team, assistance by the DCMA Commercial Item Group (CIG), and other available resources should enable the government to pay a fair and reasonable price. Industry also champions the potential time savings afforded by commercial item procurement.

In Defense acquisition, there have long been cost challenges related to "military specifications" that require solutions that differ from the commercial marketplace and therefore dramatically increase costs. The push for the utilization of commercial items to decrease costs is rooted in some core cost principles. First, commercial items allow the seller to spread overhead costs out over more customers and as a result lower prices for any individual customer who may have shouldered that burden alone. For example, a commercially developed item will reduce the need for the Department of Defense (DoD) to pay for the research, development, and testing associated with deploying a new product. Ongoing engineering requirements assuring that products are not impacted by product or part obsolescence is also spread across many buyers. Second, commercial items are controlled by free market competitive forces and as a result the need for insight into cost data is eliminated. Finally, proposing and negotiating on the basis of costs in compliance with the FAR/DFARS/CAS creates administrative burden that must be borne by the product price.

However, not every "commercial item" is created equal. The statutory definition of a commercial item allows for broad consideration of commerciality when a part is "of a type" made commercially. In this instance, competitive free-market forces do not shape prices.

There are two distinct "of a type" situations—1) where the contractor sells something themselves and 2) where they use commercial sale of an item by another party to justify their item to be "of a type." In this instance, it is not axiomatic that items previously obtained from sole source vendors using cost or pricing data (certified or not) result in savings to the government once these items are classified as commercial. As early as 1998 the Department of Defense Inspector General (DoDIG) has reported instances of overpricing when items are classified as commercial, and the buying office limited the data it considered when negotiating. There have been numerous audits where the DoDIG concluded the Contracting Officer failed to review sufficient information to determine negotiated prices were fair and reasonable for items classified as commercial.

In fact, after a commercial item determination (CID) has been made for an item previously purchased based on cost analysis, there is evidence that prices increase. Contractors assert that commercial item pricing results in a lower cost due to the removal of the administrative burden of Truth in Negotiations (TIN) and FAR Part 15 requirements. They assert that determining whether CIDs provide cost savings to the government would include other factors besides looking only at unit prices (e.g., cost to audit and provide cost analysis, maintenance of business systems, compliance with Cost Accounting Standards). However, these requirements remain for sole source acquisitions exceeding specified thresholds.

This paper explores an alternative hypothesis—that costs actually increase when products are treated as commercial—by examining prices paid for specific parts under both FAR Part 12 (commercial) and FAR Part 15 (negotiation) approaches. Using DCMA Commercial Item Group (CIG) data (not publicly available) for specific part numbers, we conducted an evaluation to determine whether the trend shows an overall increase or decrease in prices, and overall impact for the period reviewed. This deep dive into the DCMA CIG data is designed to illuminate whether there were realized cost benefits to the government.

The CIG data reviewed included 1,792 line items of data covering the period 2018 to 2024. For our review, we focused on instances where the difference between proposed and CIG recommended amounts were equal or greater than $1 million. This resulted in 87 items with a total proposed value of $1.458 billion for 28 different contractors and 66 different part numbers.

Steps we Took to Evaluate the Research Issue/Problem Statement

We determined that we would attempt to compare prices reflected in the DCMA CIG data to prices previously proposed under FAR Part 15 when cost or pricing data would have applied. To do this, we set out to contact the requestor (buying activity) for selected cases based on an analysis of the DCMA CIG data.

We analyzed the DCMA CIG data and determined that in many cases the recommended price was substantially lower than proposed. The information included prime contractor and subcontractor names and CAGE Codes, part numbers and descriptions, quantities, proposed and CIG recommended unit prices and total price, and CIG Case Numbers. The CIG data included 60 different prime contractors with 380 different part numbers with a total proposed amount of $2.714 billion. The average decrement exceeds 20% when exceptions are taken to proposed prices. We also noted that in instances where exception amounts exceeded $1 million (87 examples) the average decrement exceeded 35%. We conclude from this that the DCMA CIG group is successful in identifying overstated proposed costs when evaluating “of a type” commercial items.

The DCMA CIG data includes several useful fields, such as DCMA Case Number, Prime Contractor and Subcontractor name and CAGE codes, Description, Quantity, Proposed and Recommended Unit Price, Proposed and Recommended Price, Part Numbers (Prime and Sub), Requesting Command, End Customer, Program, and Report Date to Customer.

We determined that we needed the DCMA CIG’s assistance to identify the specific requestor associated with each Case Number. Further, in instances where DCMA Cost and Pricing (C&P) was the requestor, we needed C&P assistance to identify the originating buying activity. This delayed our research and highlights the difficulty in identifying the appropriate points of contact (POC).

Results of Data Requests to Buying Activities

We learned the following when researching the parts included on the spreadsheet provided by the DCMA CIG:

1. It is not a simple task to identify the appropriate organizations and individuals able to provide the necessary information. The identifiers we used from the spreadsheet included DCMA CIG Case Numbers, and CAGE codes for prime contractors and subcontractors. We needed assistance from the DCMA CIG to trace the Case Numbers to the original requestors (buying activities.) This was not always successful due to personnel movement, and invalid e-mail addresses. The CAGE codes were not especially useful, since they did not provide POCs for the buying activities.
2. The information provided by one buying activity was limited to the same information already presented on the DCMA CIG spreadsheet (e.g., same unit prices proposed). We requested information regarding the procurement preceding this one, and we received data on six part numbers (see Table 1).

3. One buying activity representative we spoke to stated that a recent comparison would be problematic due to the impact of COVID on pricing. This buying activity negotiates multi-year buys every five years, with the one reflected in the DCMA CIG data being in June 2022. The unit prices increased 106& from the June 2022 buy to the current buy being negotiated. These increases were due to multiple factors such as inflation, supply chain issues, and suppliers unwilling to provide multi-year pricing. Another complicating factor with this example is the subcontractor is a foreign concern.
4. Previous buys under FAR Part 15 may be at different quantities than those reflected on the DCMA CIG spreadsheet, and this along with escalation would need to be considered in any comparison.
5. Information on subcontractor proposed prices may not be readily available. The Price Negotiation Memorandum (PNM) may not have this level of detail. Further, the ability of a buying activity to readily access needed information (e.g., copies of subcontractor proposals) varies by Service.
6. We reached out to DCAA to discuss the feasibility of enlisting field audit offices with searching their files for copies of subcontractor proposals. However, this is a labor- intensive task and resource constraints prevent a detailed search of historical audit files which may not contain the desired information.
7. We utilized the commercial website nsn-now.com to research part number history to obtain solicitation numbers. We then tried to locate the solicitations for applicable part numbers in USAspending.gov, giving us the DUNS number. However, we were still unable to locate the corresponding award data within the Procurement Integrated Enterprise Environment (PIEE) or what is formerly known as the Wide Area Workflow (WAWF).
8. We surveyed the CIG for insights on procedures they use to evaluate price reasonableness.

Table 1. Responses

PRICE PAID FROM PRIME TO SUBCONTRACTOR	CIG					CUSTOMER RESPONSE				
	DATE	PART NUMBER	QTY CIG	UNIT PRICE PROPOSED CIG	UNIT PRICE RECOMMENDED CIG	DATE	PART NUMBER	QTY CUSTOMER	UNIT PRICE CUSTOMER	FOOTNOTE
DATA NOT PROVIDED						11/16/2015	8450840-1	42,000	$ 73.53	Above TINA, not commercial
	11/13/2018	8450840-1	525035	$ 67.07	$ 55.67	10/1/2018		105,007	$ 64.60	Commercial Item
						11/14/2023		63000	$ 72.38	Commercial Item
	6/23/2022	530-005378-000	75	$ 115,842.24	$ 63,939.00	NO DATE PROVIDED	530-005378-000	13	$ 115,500.00	Prior to CD Case S5113A22C0326 Date 6/23/22
								3		
								17		
								7		
								19		
								6		
								4	$ 119,320.75	
								6	$ 117,230.77	
	6/23/2022	530-005384-000	95	$ 169,296.62	$ 140,407.00		530-005384-000	7	$ 171,473.68	
								7		
								17		
								12		
								21	$ 168,428.57	
								31	$ 166,864.86	
	6/23/2022	014-000009-000	64	$ 101,385.25	$ 73,361.00		014-000009-000	16	$ 101,385.25	
								16		
								12		
								12		
								4		
								4		
	12/19/2022	19E203-2BCL97	900	$ 9,898.53	$ 8,441.40	3/15/2005	19E203-2BCL97	47	$ 3,508.00	Historical Pre Commericality
						7/26/2005		44	$ 3,530.00	
						11/23/2005		44		
						6/20/2006		90		
						8/31/2006		25	$ 5,525.00	
						1/31/2008		50	$ 5,855.00	
	10/24/2019	2060041-1	50	$ 50,309.00	$ 27,842.00	NO DATE PROVIDED	2060041	220	$ 51,181.00	Government's final negotiated per unit amount is slightly higher than the DCMA-recommended position, yet it still represents a significant cost savings compared to the initial proposed amount
	10/10/2024		220	$ 60,799.00	$ 50,366.00					

Overall Conclusions of Research

As shown in the table above, when data was provided it is not easy to ascertain the reasons for the pricing variance. Due to elapsed time, escalation could be a factor; quantity variances also impact the comparability of the data.

Our research disclosed that it is difficult to compare proposed prices for commercial parts to proposed prices before the same parts were determined to be commercial. The DoDIG audits have reported instances where historical cost-based prices for parts increased substantially when classified as commercial, but these are anecdotal examples and cannot be projected to the universe of commercial acquisitions. Likewise, our attempt to make a comparison of DCMA CIG data to prior cost-based purchase history was largely unsuccessful.

The response rate to our data request was impacted by the difficulties locating the appropriate buying office personnel. In several instances the names provided were not the current individual cognizant of the applicable program.

In one instance we were advised that the records were in paper format and would be labor and time-intensive to retrieve. Due to the time and resource constraints on our research, we did not require the additional effort.

Recommendations

To improve the ability to make meaningful comparisons in the future, consider establishing a unified part number database for government access, similar to what exists in PIEE (FED-LOG) or commercially (nsn-now.com). The database should provide historical prices paid and whether FAR Part 12 or Part 15 is applicable. The pricing data should include applicable fiscal year and quantity information, prime contractor and/or subcontractor part numbers, and National Stock Number (NSN) information for each part. For example, the first step would be to standardize an EBOM where the contractor/subcontractor provides actual prices paid in the same consistent format on all acquisitions. Next, a database for the DoD to exploit based on the standardized EBOMs could be developed. This would provide information not currently available in PIEE.

Another recommendation is for contracting officers to quickly determine if price and market-based analysis is insufficient to make a determination of fair and reasonable price. If cost data is needed and the contractor will not agree to provide, time is of the essence to elevate the matter to achieve resolution. Contractors have little incentive to provide cost data to support proposed costs for commercial items, especially if the cost data would reveal excessive profit. The excessive profit paid for commercial parts erases any perceived benefit to the government and reduces the number of items that can be procured for the warfighter.

Commercial item determinations are most often made at the subcontract level. The Government generally sees the proposed price and forms a price that they consider negotiated in negotiations with the prime, but the prime will most often negotiate a price that is different than the Government position during performance of the prime contract. Since the Government does not routinely collect and aggregate the prices paid each year, it becomes exceptionally difficult to compare pre-CID and post-CID prices.

What Goes Wrong When Negotiating Prices for Commercial Items?

Various DODIG reports have cited the following problems identified during its review of commercial acquisitions:

- A sole-source supplier with technical data rights set “market-based” catalog prices for commercial items at “what the market would bear,” and there was no competitive commercial market to ensure the reasonableness of the prices;
- Contractor refused to negotiate catalog prices for commercial items based on price analysis of previous cost-based prices, refused to provide contracting officers with “uncertified” cost or pricing data for commercial catalog items, and terminated

Government access to its cost history system;

- The contracting officer did not perform an adequate analysis when procuring sole-source commercial parts; specifically, the contracting officer used the previous DoD purchase price without performing historical price analysis and accepted the contractor's market-based pricing strategy in a noncompetitive environment without performing a sufficient sales analysis. The contracting officer did not obtain cost data to perform cost analysis;
- The contracting officer did not conduct sufficient price analysis in accordance with federal and defense acquisition regulations. Specifically, the contracting officer:
    - relied on previous over-inflated contract prices to determine the current contract prices;
    - did not sufficiently analyze the "commercial of a type" parts to determine whether the sales of comparable parts supported the contract prices;
    - accepted excessive prices for new quantity ranges; and
    - did not compare commercial sales to Government sales to determine whether sales were sufficient to support commercial part prices;
- The contracting officer did not appropriately determine fair and reasonable prices for sole-source commercial spare parts purchased from the contractor. This occurred because the contracting officer did not conduct a sufficient price analysis. Specifically, the contracting officer:
    - relied on sales data that did not include customer names;
    - did not review commercial sales quantities; and
    - accepted prices for sole-source commercial parts with no commercial sales.
    - Further, the contracting officer did not question the commercial off-the-shelf classification for parts with no commercial sales, and did not require the contractor to comply with a contract requirement to submit negotiation documentation within stated timelines.

In two of the DoDIG reports reviewed, the DoDIG had to issue subpoenas to the contractor to obtain other than certified cost data. The contracting officer is at a decided disadvantage because contractors are hesitant to provide cost data to support pricing for commercial items. Cost data is last on the list of items the contracting officer should review to determine price reasonableness.

Conclusion

In conclusion, identifying potential savings (or, alternatively, cost increases) for "of-a-type" commercial supplies and services is not straightforward, commerciality savings is not as clearly defined in procurement of government supplies and services based on the currently available information. This is due to several factors: 1) EBOMS are not standardized to include FAR Part 12 or 15 applicability making comparisons difficult; 2) the supplies or services are often provided by subcontractors. The government may complete negotiations with the prime before the prime completes negotiations with the subcontractor. In these cases, only the prime contractor has visibility of the negotiated price between the prime and subcontractor for a particular supply or service. If EBOMS were standardized and required to include prices negotiated between prime and subcontractor, a database could be developed and used by the government to evaluate fair and reasonable pricing in future buys.

In the absence of a true commercial market to determine a fair and reasonable price for an "of-a- type" commercial item, there must be better tools available to the contracting officer. While cost data to support proposed commercial prices is a "last-resort," it may be the only valid way if other methods have failed. Highly redacted or limited sales history, for example, is not sufficient just because the contractor/subcontractor makes proprietary assertions or has limited sales. If the contracting officer requires cost data to establish a fair and reasonable price, the regulation should make it easier to obtain cooperation. The DCMA CIG data used in our research clearly shows that proposed prices for commercial "of-a-type" items are often overstated. Contracting officers should continue to seek assistance from the DCMA CIG for pricing help on commercial items.

References

DoD. (2019). *2019 Department of Defense guidebook for acquiring commercial items: Part A commercial item determination*. https://www.acq.osd.mil/dpap/dars/pgi/docs/DoD_Guidebook_PartA_Commercial_Item_Determination_07_10_19.pdf

DoDIG. (1998, June 24). Commercial and noncommercial sole-source items procured on contract N000383-93-G-M111 (Report No. 98-064). Department of Defense. extension://efaidnbmnnnibpcajpcglclefindmkaj/https://media.defense.gov/1998/Feb/06/2001713 743/-1/-1/1/98-064-s.pdf

DoDIG. (2014, July 3). *Defense logistics agency aviation potentially overpaid Bell Helicopter for sole-source commercial spare parts* (Report No. DODIG-2014-088). extension://efaidnbmnnnibpcajpcglclefindmkaj/https://media.defense.gov/2018/Jul/24/20019461 92/-1/-1/1/DODIG-2014-088.PDF

DoDIG. (2015, May 8). *Defense logistics agency did not obtain fair and reasonable prices from Meggitt Aircraft Braking Systems for sole-source commercial spare parts* (Report No. DODIG-2015-120). extension://efaidnbmnnnibpcajpcglclefindmkaj/https://media.defense.gov/2015/May/08/2001713 503/-1/-1/1/DODIG-2015-120.pdf

DoDIG. (2016, February 16). Defense logistics agency did not appropriately determine fair and reasonable prices for F108 engine sole-source commercial parts (Report No. DODIG-2016-047). extension://efaidnbmnnnibpcajpcglclefindmkaj/https://media.defense.gov/2016/Sep/08/2001774 201/-1/-1/1/DODIG-2016-047.pdf

A Centralized Financial Reporting System for Improved Data Access and Analysis

Kumiko Dunn, PhD—is a Data Scientist at the Naval Undersea Warfare Center (NUWC), Division Keyport. [Kumiko.T.Dunn.civ@us.navy.mil]

Karen Knewtson—is a Program Analyst at the Naval Undersea Warfare Center (NUWC), Division Keyport. [Karen.L.Knewtson.civ@us.navy.mil]

Emmanuel Delgado—is an IT Specialist at the Naval Undersea Warfare Center (NUWC), Division Keyport. [Emmanuel.E.Delgado.civ@us.navy.mil]

Matthew Bauchspies—is a Data Scientist at the Naval Undersea Warfare Center (NUWC), Division Keyport. [Matthew.T.Bauchspies.civ@us.navy.mil]

Dallas Rosson, DEng—is a Principal Engineer at the Naval Undersea Warfare Center (NUWC), Division Keyport. [Dallas.J.Rosson.civ@us.navy.mil]

Abstract

Throughout the Department of the Navy, accurate and timely understanding of the command's business financials is critical for decision making. The inability to directly connect data analytics applications to financial data sources and the availability of multiple sources for data results in large variability in the information reported and significant time to download and pre-process the data.

The Naval Undersea Warfare Center, Keyport Division developed the Centralized Financial Reporting (CFR) system that allows analysts to access multiple datasets required for financial management. The cleaned and validated datasets are available via SharePoint and a Power BI Premium workspace. Ready-to-use files, dataflows, and semantic models stored in these workspaces allow for immediate visualization and data analysis, resulting in a significant time savings and an estimated cost savings of more than a million dollars on these tasks. Additionally, easy accessibility to transaction-level details allows a quick evaluation of the contributors to the financial health of Keyport business.

In this paper, the CFR system is described and evaluated, including the three concerns it's use mitigates: the significant resources required for financial reporting, difficulty evaluating factors affecting information garnered from the data, and discrepancies in reporting that are experienced with the current methods used to evaluate business financials.

Introduction

As the Department of the Navy (DoN) works towards being both accountable and transparent in financial matters, limitations in some of the financial systems have hindered the meeting of some financial reporting goals (DoN, 2024). As would be expected, similar difficulties persist across the warfare centers (WFC).

At the Naval Undersea Warfare Center (NUWC), Keyport Division, direct connections to business financial data via data analytics applications do not exist, resulting in the need to download, clean, pre-process, and validate the data from multiple data sources on weekly or monthly bases. This significant burden on the analysts reduces their ability to focus on in-depth analyses of the data or requires them to work outside of their schedule hours.

In attempts to reduce this burden, summary tables are often pulled from the data sources. As would be expected, the use of summary tables restricts the ability to delve into the transaction-level data. Without access to the transaction level data within a dashboard, additional work is required to further evaluate the data for any reason (e.g., ad hoc data calls that are not answered with the summation tables or errors in the summation tables).

The last concern resulting from the lack of a direct connection to the data is that additional data sources have been created that pull data from the authoritative data sources to make access to the data easier (e.g., business financial transactional data is accessed through DoD Resource Planning [DRP], DoD Data Warehouses [DDW], and Jupiter [DoN, n.d.-a]). This results in potential discrepancies between reports and dashboards created with the different data sources. The reasons for the discrepancies are varied but may result from the timing of the data being made available or errors contained in one data source but not in another.

To mitigate data access concerns across the Navy, the DoD and DoN have created Advana Jupiter, an enterprise data and analytics environment (Booz Allen Hamilton, n.d.; DoD, n.d.; DoN, n.d.-b). Currently, Jupiter is working to provide users access to DRP data through both the analytics applications housed within Jupiter and via outbound connections (DoN, n.d.-a; DoN Chief Information Officer, 2020). At the time of this writing, the DoD has paused development on Advana, which is the DoD environment that houses Jupiter (Williams, 2024). Once it has been made accessible, it is anticipated that this will be a preferred method of accessing DRP data. Until then and the time when all required datasets are available in Jupiter, the Centralized Financial Reporting (CFR) system will make data accessible at NUWC Keyport. This system is described below.

Objective and Approach

Datasets

NUWC Keyport Program and Financial Analysts rely on several data sources to report the financial health of the projects they monitor. As a Working Capital Fund (WCF), the data sources of interest include DRP, DDW, DoD Planning System (DPS), Enterprise Quoting System (EQS), DoD Work Management System (DWMS), DoD Investment Reporting System (DIRS), and Excel spreadsheets on individual analysts' computers.

DRP is the authoritative data source for every charge that has been applied (i.e., transactions) or is anticipated (commitments and obligations) to be applied. From DRP, several reports can be obtained. Two commonly used reports are the Actual Transactions table and Actuals Obligations table. The Actual Transactions table details every transaction, whereas the Actual Obligations table details the commitments and obligations. Although DRP is the authoritative data source for financial transactions, the data contained within it can be difficult to parse out. Therefore, some analysts will choose to get data from DDW because it provides a cleaned and processed view of the data. DDW can also be used to obtain categorical data that will provide analysts more information regarding the information contained in the Actual Transactions table or Actual Obligations table. For example, the Labor Charge Category table out of DDW provides more information about the labor charges (e.g., breaking out overtime and regular labor data) but the Actual Transactions table does not. Additional tables that can be pulled from DDW include the Network Categories table that provides categorical data for each network like the network title and center code associated with network and the Sales Category table that provides information about each sales order (e.g., the network associated with a sale, the purchaser, date of purchase, etc.).

In the budget planning arena, DPS, DQS, WMS, and DIRS are used by different teams at NUWC Keyport based on factors such as the customer and product being produced. Therefore, there is no single authoritative data source for the budget planning. An additional consideration is that some monthly phase plans have historically been documented in Excel and sent by email for aggregation and visualization of the planned and actuals. Therefore, to obtain the same information for each of the projects reported in the Power BI visuals, the CFR system includes Microsoft Power Apps for analysts to enter their monthly phase plans. The Power Apps provide a method for easy project planning data entry into the CFR system. The apps also allow the analysts to make corrections and updates to plans, and to allow an “at a glance” view of a particular plan without accessing a second application.

The last set of data is obtained from the Finance office, including static factors, cost elements, work centers, and allocations tables. The static factors table provides the analyst with daily, monthly, and yearly anticipated workyear (WY) factors. By using these factors, the WYs for each timeframe can be calculated by dividing the hours by the factor. The WYs are split up by regular and overtime (OT) charges. The cost element table provides categorical data about cost elements (or items that were purchased), e.g., title and category of the cost element. The work center table provides information about the department, division, and branch associated with the work center. The allocations table provides information about the funds budgeted for overhead and service funds.

Data Access and Storage

Data access options were investigated. As stated above, direct connection to the required data sources from our system is the preferred method; however, none of the data sources listed above currently allow a direct connection. Jupiter provides a possible indirect access solution. As we work to obtaining access to this connection, we opted to proceed by downloading the data to a location that could be accessed across command.

Several storage solutions were investigated including cloud storage options (e.g., Azure), SharePoint, and on prem databases. SharePoint was chosen for the reasons outlined below.

1. The relatively small size of the datasets (< 10 GB).
2. The prohibitive time and financial requirements for the procurement of cloud storage or database solutions.
3. The readily available and “freely” accessible SharePoint storage via Flank Speed.
4. SharePoint being a Microsoft solution with connectors available for easy connectivity to other Microsoft applications (i.e., Power BI, Power Automate, and Power Apps).

Having determined the required datasets and the storage locations, we downloaded, validated, and uploaded the data into SharePoint. The data was obtained from the sources indicated in Table 1.

Table 1. Datasets and Data Sources

Dataset	Data Source
Actual Transactions	DRP
Actual Obligations	DRP
Financial Plans	Analysts via MS Power App/SharePoint
Network Categories	DDW
Labor Charge Categories	DDW
Sales Categories	DDW
Workyear Factors	Finance
Purchase Categories	Finance
Work Center Categories	Finance
DOD Workyear and Financial Allocations	Finance
Investment Planning	DIRS

Power BI Dataflow

After the data was stored in SharePoint, Power BI dataflows were created. Power BI dataflows allow for entire datasets to be cleaned, processed and made available for use in Power BI desktop or Power BI online. It also provides increase security by preventing access to the underlying data (Microsoft Learn, 2024b). It is important to note that the dataflows are available on the Power BI premium services (Microsoft Learn, 2024b).

To date, seven dataflows were created that include the data housed in SharePoint. The dataflows are as follows: Actual Transactions table, Common Dimension Tables, Dates and Factors, Net Operating Result (NOR) analysis, Phase Plans, and WKYRs. The Actual Transactions table dataflow consists of the aggregate of all Actual Transactions tables pulled from first use of DRP to current. The Common Dimension Tables dataflow consists of the Network Categories, Sales Categories, Purchase Categories, Work Center Categories, DoD WYs and Financial Allocations, and miscellaneous crosswalk tables. Dates and Factors includes a dates table (i.e., a table of all dates from 2012 to 2030) and WY factors table. The NOR dataflow contains miscellaneous tables related to the Commands ability to hit the NOR target by the end of the fiscal year (FY). The Phase Plan dataflow contain the aggregated phase planning tables for all FY starting in FY25. The WKYRs dataflow contains the aggregated Labor Charge Categories tables from first use of DRP to present. The dataflows will be augmented with additional datasets as categorical data for different funding types are obtained.

Power BI Semantic Model

Using the Power BI dataflows, two Power BI semantic model were created (investments and overhead/service) and a third one will be created in FY25 for direct funds. A Power BI semantic model is a set of one or more tables that have been cleaned and related in the Model view of Power BI desktop (Microsoft Learn, 2024a). It provides the user access to the cleaned and preprocessed data with defined relationships between tables, allowing users to focus on visualizations and report creation. The semantics models are made available via the Power BI Premium service. An example of a Power BI Semantic model is illustrated in Figure 1, which shows the Overhead and Service model used in the CFR system.

Figure 1. Illustration of the Overhead and Service Semantic Model in Power BI

These semantic models can be downloaded from the Power BI Premium Service workspace to another Power BI workspace or an individual's OneDrive workspace. The download provides access to both the semantic model and any visualizations attached to the model, allowing the user to create new or modify existing visualizations.

Power BI Visualizations

Using the CFR system, all the data described above is accessible by analysts across command to analyze and create reports. The SharePoint Lists and Excel workbooks allow the use of any application to analyze and create visualizations of the data. Using the Power BI dataflows and/or the Power BI semantic data model, analysts can connect to the data via Power BI Desktop or Power BI online.

Currently, Keyport analysts have been working with the Power BI semantic model and have created many visualizations of the Overhead, Service and Investment financials. To illustrate the capabilities of the system, the Overhead and Service visualizations that are currently available in the command-level Tableau (Tableau Software, 2024) dashboards were recreated in Power BI. Further, to illustrate the power of making transaction-level data available, additional reports were created that allow an analyst to drill into the data to answer more specific questions regarding the health of their portfolios (e.g., Overtime Analysis, Service Revenue Analysis, drill-throughs into each transaction, etc.).

Previous, NUWC Keyport command-level reports have been created using Tableau software (Version 2024.2.2). Figure 2 illustrates the existing command dashboard while Figure 3 illustrates that the CFR system houses a similar visualization. Figure 5 illustrates the data that is not available with the Tableau visuals but can be quickly extracted from the CFR system.

Figure 2. Tableau Dashboard that is Currently Being Used on Station Contains a Bar Chart, a Line Chart, a Bar/Line Combination Chart, and Aggregated Financial Information

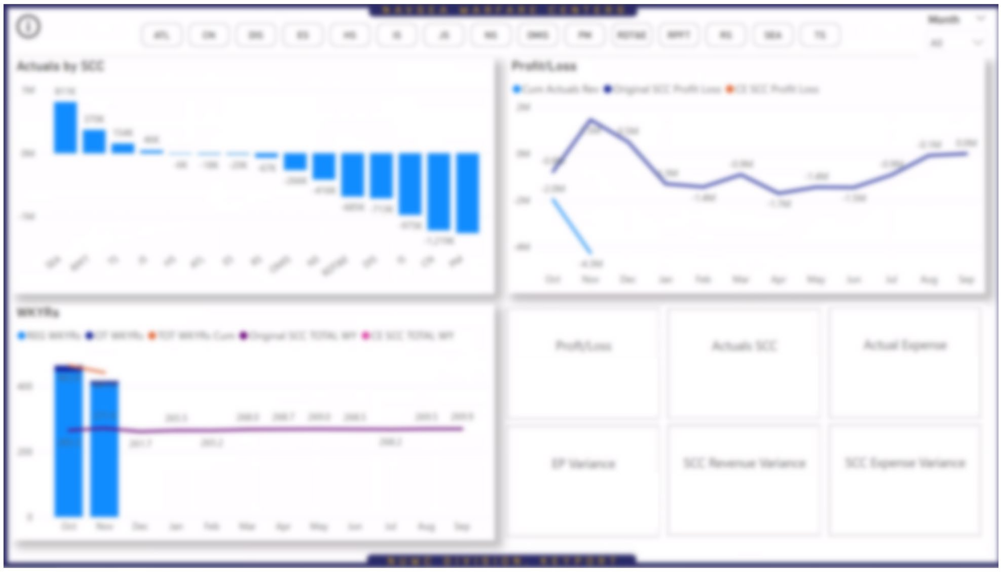


Figure 3. Service Analysis Power BI Report Illustrates the Ability to Mimic the Existing Tableau Dashboards

The Service Revenue Analysis Power BI Report (Figure 4) illustrates the granularity that can be provided. This report provides the user with Service specific revenue by type, by the Purchase Category name, and by the department providing the revenue. Previously, this information was not readily available to the line analyst. Instead, the analyst would have to take significant time to comb through Excel spreadsheets to gather the data illustrated in Figure 4. Then, they would have to find the crosswalk that would link the network to the owner of the funds. For an experienced analyst, this could take several hours. For a new analyst, they might not know how to get this information. With CFR, the analyst no longer has to do the extra legwork. They can drill into the data, as needed.

Figure 4. Service Revenue Analysis Power BI Report Provides the Ability to Delve into the Data in a Way that is Not Currently Available in the Tableau Dashboard

The Revenue Analysis (Figure 5) shows how you can now drill into the data to extract important information that was not available previously. In this example, the amount paid to a Service fund for a specific cost element can be determined per department on station. Although this shows information at the departmental level, the information can also be shown at the division, branch, and even network level.

Figure 5. Revenue Analysis Illustrates the Ability to Drill into the Data

CFR Inputs and Outputs

Up to now, we have discussed the various components of the system. Figures 6 and 7 illustrate the entire CFR system. Figure 6 illustrates the inputs to the CFR system including the data input by the analysts on project phase planning, the data obtained from DRP, DDW, the Finance office, and DIRS. This data is then stored in SharePoint either as a List or as an Excel Workbook in a SharePoint folder. Using this data, Power BI dataflows are created, then using the dataflows, the Power BI semantic data models are created. It is important to note that the analysts have access to the data at each of the four input stages. This allows them the freedom to use other data analytics and visualizations programs other than Power BI, as well as the ability to choose their preferred way of accessing the data. In Figure 7, four different CFR output options are illustrated: in-depth data analytics, customized Power BI reports, command-level Power BI reports, and Power BI Apps that allow for aggregation of several different Power BI report sets. The last option provides stakeholders and leadership to access several different reports and dashboards via a single URL. This would be beneficial for command-level leadership, who keep track of many different financial reports across the command.

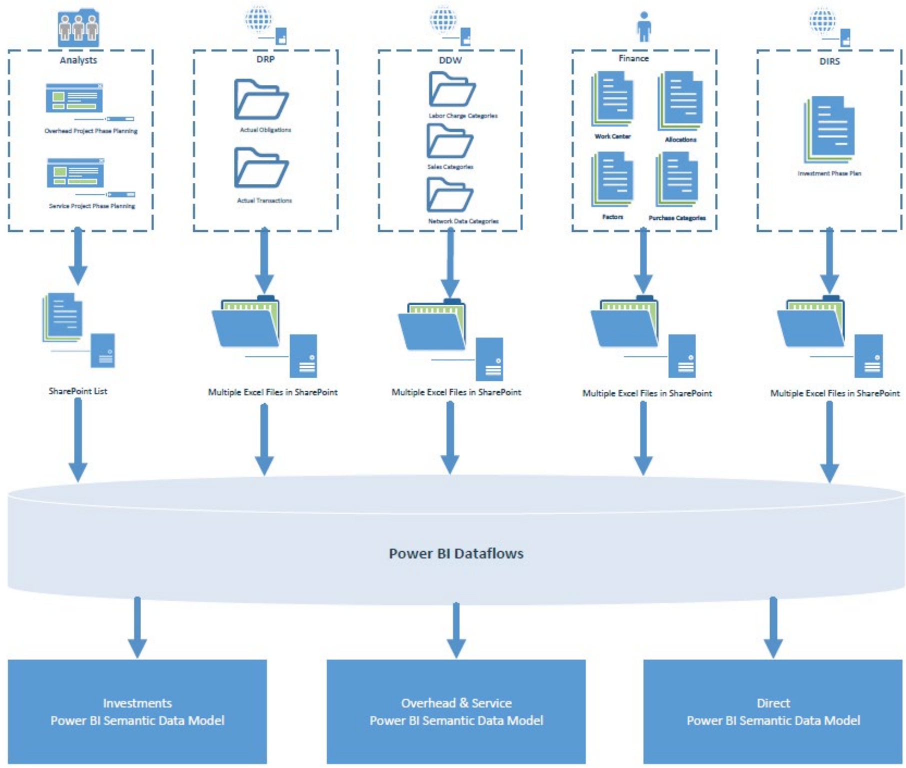


Figure 6. CFR Data Inputs and Processing

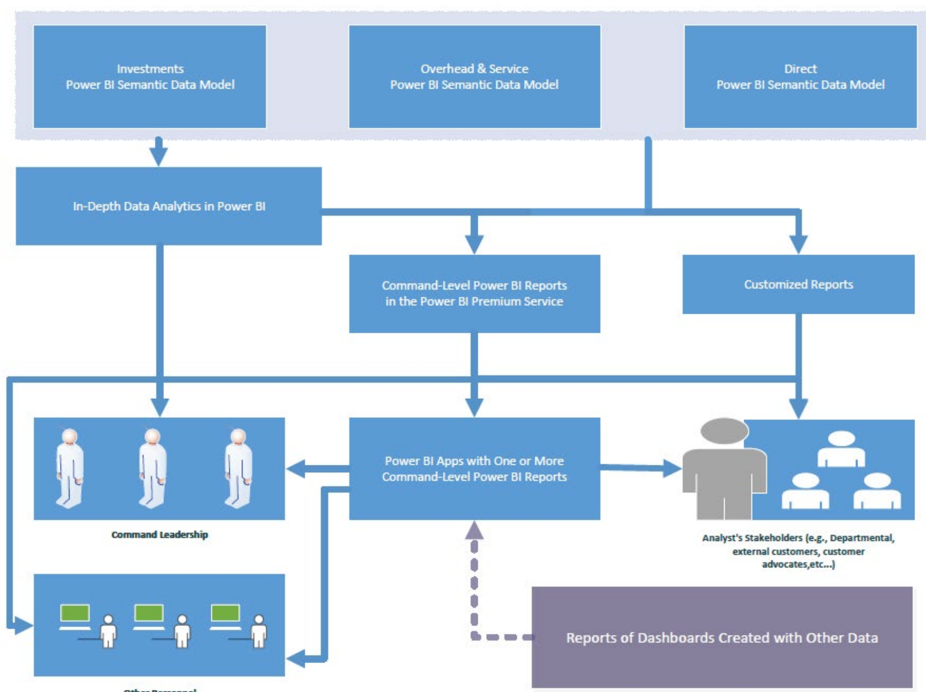


Figure 7. CFR Data Analytics and Power BI Reporting Options

Key Insights

Use of the CFR system has the potential to improve financial reporting across the NUWC Division Keyport and potentially across other warfare centers. These improvements come in the form of a reduction in time and resources for compiling financial information, an increase in the level of detail that can be obtained from the data, and reduction in discrepancies in reporting. Although not yet discussed, it is important to note that to achieve these benefits, training is required. As such, a training program is being developed as outlined in the training section.

Reduction in Time and Resources

Financial reporting at NUWC Keyport requires the analyst to obtain data from sources of record, clean the data, then analyze and create any required reports. The time to complete this varies from analyst to analyst based on the number of data sources, the amount of data, the analyst's technical expertise in automated data cleaning options, and the number of ad hoc data calls the analyst receives. For periodic reporting, this might take hours to days for OH or Service monthly reporting, and eight hours per week for direct weekly reporting. The ad hoc data calls can add a significant amount of time to their monthly or weekly reporting time because of the need to comb through the data to get the necessary information. It also requires a level of data expertise that newer analysts may not possess, resulting in the inability to provide the required information.

A significant benefit of the current system is the reduction in time needed to create the periodic and ad hoc reports. This is accomplished by eliminating the need to access and clean the data. To quantify this time saving, the Pareto principal (Pragmatic Editorial Team, 2024) can be used. In data science, it is estimated that it takes 80% of the time to clean and prepare the data and it takes 20% of the time to conduct the analysis. With this in mind, the new system has the potential to reduce report creation time by roughly 80%.

If every analyst on station puts in 8 hours/week for data preparation and analysis. The current system would reduce this time by 6.4 hours/week. To estimate the cost saving in labor dollars, a few assumptions will be made.

1. An analyst costs $158/hour.
2. There are 35 analysts.
3. There are 48 work weeks/year.

Based on these assumptions, this would result in a savings of almost $1.7 million/year on analysts work related to cleaning and pre-processing data. The time savings would allow analysts to focus on improving their data analytics skills, providing better insights of the data, training new analysts, and addressing their work backlog. It also has the potential to improve the analysts' morale because of the reduction in overtime needed and by providing a sense of being able to successfully accomplish their tasking.

It is important to note that these are estimates are based on the assumptions that have been outlined. Several factors may affect these calculations, including the need to conduct additional analyses (e.g., some analysts may need to add a Data Analysis Expression [DAX] measure or two to obtain the information requested by their stakeholders). Additionally, an individual will need to pull the data into the system and make sure the automated processing has been completed accurately. The weekly data pulls and processing takes roughly 3–5 hours/week. This cost would have to be added back into the above estimates. This would add roughly $40,000 back into the above estimates. In light of the $1.7 million in estimated savings per year, this is negligible.

Increased Granularity in Available Data and Reporting

Increased granularity in the data is provided by making the transaction-level data available. Currently used reporting mechanisms do not allow for deep dives into the data, resulting from the use of summary tables or other methods used to create the reports. Unfortunately, this method obscures the underlying data.

By making the transaction-level data and dimension tables available, the information can be presented in a way that is more comprehensible. For example, the unique identifier of a Purchase Category is a number that does not convey what that purchase category is unless a crosswalk is available that links the number to the name of the Purchase Category. In the current system, we have provided a semantic model that links the Purchase Category dimension table to the tables with transaction-level data. This allows the analyst to quickly create reports that provides information that is easier to consume without needing to pull the additional table. A secondary benefit is that a newer analysts can use this comprehensive dataset to learn how to answer some of the more complicated data calls that previously required a detailed knowledge of the datasets, where the data was stored, how to access the data, and what questions the datasets could help answer.

Finally, the use of transaction level data allows the analyst to quickly evaluate any errors in the data or inconsistencies between two reports. Within Power BI reports, it is possible to create pages that allow the analyst to drill into the data in a visualization to see the underlying data that is included in the aggregation. This can then provide the analysts with information about what is included in the aggregation, and potentially what data is producing the erroneous result(s).

Reduction in Report Discrepancies

Discrepancies in reporting have previously caused significant time to be spent to determine the source of the discrepancies. Common sources were the use of different data sources and the use of summary tables. Copying and pasting the data incorrectly into tables was another source.

To reduce the discrepancies, the CFR system provides centralized access to the data and semantic data model. The semantic data model pre-links the transaction-level data to dimension tables used to provide categorical information that allows for easier comprehension of the data. This allows the analysts quicker access to the data, improving the chance they will use the same validated data. Additionally, the use of the semantic model with defined table relationships provides the less experienced analysts with the ability to start creating visualizations without an in-depth understanding of the required relationships.

Training

As we strive to reduce the burden of accessing and pre-processing data, we have created a system that uses Power BI dataflows, semantic models, and reports. Although the system reduces the pre-analytic processing, it also incorporates technology that is relatively new to the analysts on station and introduces complexities of its own (e.g., accessing a semantic model or connecting to a Power BI dataflow on the Power BI premium service). To help analysts learn the new system and to encourage its use, training sessions are being developed. This training will be provided as a part of an existing Power BI training series. The training consists of basic information on how to use Power BI Desktop and Power Query, as well as information about the data housed in the Power BI dataflows, information about the connections and how best to use them, and hands-on instructions on connecting to the CFR data, using the provided visuals and creating new ones.

Limitations

While the CFR system addresses several of the hurdles that analysts at Keyport traverse, not all concerns have been addressed. As stated in the introduction, direct access to all the datasets would be the ideal option. This would reduce the need of someone downloading the data. Instead, a Power BI dataflow and semantic model could be created from a direct connection to the data at the source (e.g., by connecting to DRP instead of to the files downloaded from DRP). Additionally, it will reduce the possibility of any errors introduced in the download step.

Conclusions

Management of the financial portfolio of a WCF is time consuming and costly because of the complexities of this type of fund. The NUWC Keyport CFR system was developed to significantly reduce the time and cost of managing these portfolios, by providing analyst access to all relevant data from a central location. This data is validated and preprocessed and is also directly accessible via Power BI for reporting purposes. By making this preprocessed, comprehensive dataset easily accessible, Keyport analysts can spend more time on in-depth analyses and report creation, providing their stakeholders the reports necessary to make well-informed and sometimes critical decisions.

References

Booz Allen Hamilton. (n.d.). *Advance enterprise analytics at the defense department*. https://www.boozallen.com/d/insight/thought-leadership/advanced-enterprise-analytics-at-the-defense-department.html

DoD. (n.d.). *Advana*. https://advana.data.mil/

Department of the Navy. (n.d.-a). *Jupiter*. https://jupiter.data.mil/

Department of the Navy. (n.d.-b). *What is Jupiter?* https://www.dau.edu/sites/default/files/webform/documents/26876/What%20is%20Jupiter.pdf

Department of the Navy. (2024). *United States Department of the Navy agency financial report: Fiscal year 2024*. https://www.secnav.navy.mil/fmc/Documents/FY%202024%20DON%20AFR.pdf

Department of the Navy Chief Information Officer. (2020). *Jupiter: Bringing the power of data analytics to the DON*. Department of the Navy. https://www.doncio.navy.mil/mobile/ContentView.aspx?ID=13804&TypeID=21

Microsoft Learn Challenge. (2024a, August 28). *Semantic models in the Power BI. https://learn.microsoft.com/en-us/power-bi/connect-data/service-datasets-understand*

Microsoft Learn Challenge. (2024a, September 25). *Introduction to dataflows and self-service data prep—Power BI.* https://learn.microsoft.com/en-us/power-bi/transform-model/dataflows/dataflows-introduction-self-service

Pragmatic Editorial Team. (2024). *Overcoming the 80/20 rule in data science*. Pragmantic Institute. https://www.pragmaticinstitute.com/resources/articles/data/overcoming-the-80-20-rule-in-data-science/

Tableau Software. (2024). *Tableau* (Version 2024.2.2) [Computer software].

Williams, L. C. (2024). *Pentagon pauses Development of its go-to data analytics tool*. Defense One. https://www.defenseone.com/defense-systems/2024/06/pentagons-go-data-analytics-platform-under-construction/397255/

Augmenting Intelligence: Acquiring Trustworthy Technology

Carol J. Smith—is a Principal Research Scientist at the CMU Software Engineering Institute in the AI Division. Smith has been conducting research to improve experiences with complex systems for over 20 years. Smith trailblazed human-computer interaction (HCI) research for AI, autonomous vehicles, and other emerging technologies and is recognized globally as an advocate for human-centered AI. Smith is a co-author of the DoD Defense Innovation Unit Responsible AI Guidelines and has presented over 250 talks and workshops around the world. She teaches at CMU and holds an MS in HCI from DePaul University. [cjsmith@sei.cmu.edu]

Abstract

Procuring understandable systems that improve human effectiveness and reduce (or at least maintain) overall risk has become even more complex when considering the acquisition of artificial intelligence (AI) systems. Current acquisition guidance provides a strong foundation but is not sufficient to identify and effectively assess emerging technologies such as generative artificial intelligence. These systems can provide value and be exceedingly helpful in the right situation. This paper provides actionable guidance to Navy acquisition teams so that they can quickly and effectively identify and procure the best AI systems and reduce risks associated with these major investments.

Introduction

This paper provides guidance for acquisition teams to quickly select, test, and implement AI systems with a streamlined practice that informs procurement teams, enables operational experimentation, and ensures the continuity of capabilities. Successful acquisition and adoption of emerging technologies requires an ability to identify aspects that will enable their trustworthiness. This work builds on the Software Acquisition Pathway (Defense Acquisition University [DAU], 2020) which the recent Department of Defense (DoD) memo (Hegseth, 2025) directs all DoD components to adopt. In addition, the paper incorporates decades of research and experience developing and designing complex and dynamic systems that work with, and for, humans.

The approaches described in this paper will enable the United States to gain an early advantage by quickly identifying the best AI solutions to achieve our goals, that meet the needs of the workforce, and that reduce cost and risk. The introduction of generative AI systems has ignited a significant leap in awareness of the capabilities of AI and will be specifically addressed in the paper. This guidance is designed for use cases that involve some risk (forecasting, planning, anomaly detection, etc.). This guidance can also be used for very low risk systems (e.g., movie recommenders), but it is not appropriate for extremely high-risk systems such as robotics or weapons systems. This guidance assumes that the organization is making a significant investment in the new system and is interested in assuring it is adopted effectively.

Informed Approach

Successful acquisition starts with a brief assessment to determine the organizations' readiness for AI technology and that AI is a match with their needs. While there are many formal methodologies for requirements gathering, this approach is focused on gaining relevant knowledge and informing the acquisition process. There are two areas of focus, the first being to identify relevant needs and constraints, and the second to identify resources and capabilities.

Relevant Needs and Constraints

Briefly analyzing the current system and the context of use is extremely important to ensure the right product is identified. The following questions will guide the team to focus their efforts:

- What is the problem that needs to be solved, and for whom?
- What is the existing level of risk in the current context of use?
- What are the perceptions of AI in the organization? What about the expected end-users' perceptions?
- What are the potential impacts of a new system – both beneficial and harmful?

Effective adoption of any type of software, including AI systems, requires basic knowledge of the existing context of use and the people who will use the system and then matching this with a system that provides value. "Business value continues to be a challenge for organizations when it comes to AI," according to Leinar Ramos, senior director analyst at Gartner (2024). The current processes should be reviewed to identify what is working well and what is not. That information forms a baseline of performance that acquisition teams can use to make informed decisions about potential systems. This effort may also lead to identifying areas that may need more consideration.

Individual perceptions of AI can be the most critical factor in how successful a new AI system will be. End-users who have extremely high expectations for the system may determine that oversight is not necessary. Overtrust of an AI system can result in the system being used for tasks it was not designed to do. Failures due to overtrust can be simply frustrating or, at their worst, can lead to disastrous situations such as described in Dastin (2018), Smiley (2022), and Armstrong (2023).

Under some conditions, AI tools may in fact limit, rather than enhance, scientific understanding (Messeri & Crockett, 2024). For example, scientists using AI tools for research may falsely believe they are exploring a space of all testable hypotheses, whereas they are actually exploring a narrower space of hypotheses that are testable with AI tools (Messeri & Crockett, 2024). Or they could become vulnerable to an illusion of objectivity, in which they falsely believe that AI tools do not have a standpoint or are able to represent all possible standpoints (Messeri & Crockett, 2024).

The addition of an AI system can supercharge an organization and significantly augment individual productivity. Along with these benefits, the dynamic nature of AI increases the level of risk that must be accepted by those using the system. An AI system can also increase risk for those affected by decisions made with or by the system. A system that adds more risk or requires additional fact checking may not be appropriate in contexts when decisions need to be made quickly or when correct outputs are required.

The organization's norms and culture are an important aspect of successful adoption. Engaging end-users and the team that will manage the system in a brief brainstorming activity to consider "What Could Go Wrong?" (Martelaro & Ju, 2020) will support the identification of risk for the system, increase understanding of the context of use, and can be a method to reduce fear by exploring difficult topics. User experience (UX) activities such as "Black Mirror" Episodes (C. Fiesler, personal communication, 2018) and Abusability Testing (D. Brown, personal communication, 2019) can also support these goals. Each of these methods entails minimal effort and will make a positive impact on the quality of the system and its adoption. As a reminder, these methods are not sufficient for high-risk system acquisition.

Resources and Capabilities

A brief analysis of resources such as data and staffing will also support the identification of the right system. Use the following questions to guide the team:

- How representative is the training dataset to the intended operational context?
- What experience does the organization have managing complex systems?

- What resources are available (or needed) for monitoring and managing the AI system over the expected period of operation?
- What type of AI system(s) are a good match with the problem to be solved?

An AI system is most effective when it is trained on data that fits the use case. A quick review of the data the team intends to use will be helpful in preparing for an AI system. For example, at the SEI, we quickly found that a computer vision system trained to identify tanks in a lush green location was not useful in identifying tanks in a desert by doing a relatively small experiment. Exploratory data analysis (U.S. Environmental Protection Agency, 2025) combined with qualitative methods can support data understanding. There are also sources of guidance from the DoD and others to support a data-driven culture, such as Gebru et al. (2021), Defense Information Systems Agency (DISA, 2025), and DrivenData (n.d.).

Any data that is used for training or otherwise contained within the system has the potential to be obtained through use of the system. Guardrails and other precautions are helpful and will work in most situations, but if undesirable or harmful information, personally identifiable information, or other types of non-public data are potentially in the system, the team will need to accept the risk of exposing that data. This is a particular risk with systems using generative AI which are designed to generate new combinations of data. These systems provide many opportunities and benefits and can be fine-tuned with additional information about a specific topic; however, they do not reliably retrieve specific data, nor are they likely to successfully reproduce the same outputs. Additionally, many claims made today about what generative AI can do are overhyped (Carleton et al., 2025).

Teams will be most successful integrating and managing AI systems when they have previous experience managing complex systems, strong technical capabilities, and a desire to learn. Cybersecurity is unlikely to be affected directly, but nearly all other aspects of the existing systems, applications, and integrations will likely be affected. Similarly, the teams' preparation for monitoring and managing the AI system will influence the systems successful adoption. "As organizations scale AI, they need to consider the total cost of ownership of their projects, as well as the wide spectrum of benefits beyond productivity improvement," said Ramos (Gartner, 2024).

As with any software system, as previously mentioned, the acquisition team needs to understand the use case and be provided with clear criteria for purchase selection. help the procurement team narrow the choices and assess products for potential suitability for its intended purpose. The new system should improve the situation and perform at least as well as the previous system.

Selecting an appropriate AI system for the problem to be solved can be a challenge, as there are many types of systems and each has strengths and weaknesses. For example, generative AI systems such as large language models (LLMs) are very popular currently, but they are not the right choice for every situation (Tao, 2024). An LLM can be an excellent solution to meet the needs for a chatbot or generating content, but it is not a good solution for decision intelligence or forecasting which require tools that can retrieve specific information or analyze data. A well-defined problem to solve will enable easier matching to an AI solution.

Operational Experimentation

Once the initial vendors and AI system selections are identified, it is time to validate suitability, and the only way to get UX design right is to test it (Moran, 2024). Operational experiments can be conducted with a full AI system, a minimum viable product (McDonald, 2023), or a clickable (low code) prototype. The AI system does not need to be fully functional or fully integrated into the environment, but it does need to at least provide an understandable and representative experience of the primary tasks it is intended for. This can be challenging for vendors but is a reasonable request for a substantial investment.

The people who will use the product (end-users such as warfighters, analysts, operators, etc.) need to be given access to the system or prototype in their typical environment. With just a short introduction to the system, the end-users should be able to use the system to do the core tasks it is expected to support. They should be able to interpret the output of the AI product and be able to determine if the system is working as expected. This activity is akin to usability testing (Moran, 2024), and if the system is well designed, the end-users should need only minimal support and not specific direction.

This part of the assessment is subjective by design to enable identification of failures that will erode trustworthiness as early as possible. End-users are typically the ones to discover AI technology failures, and those negative experiences are risk indicators of deteriorating trustworthiness (Gardner et al., 2023). Organizations employing these systems must therefore ensure that end-users are supported with:

- indicators within the system when it is not functioning as expected
- ability to report when the system is deteriorating or not operating properly
- information to align their expectations and needs with the potential risk the system introduces (Gardner et al., 2023)

Before determining whether to employ a new AI technology, ask these questions (Gardner et al., 2023):

- What are the limitations of the system's functionality?
- What are the safety controls to prevent this system from causing damage? How can these controls be tested?
- Is the development team able to understand and audit the output of the tool?
- How was the model trained? Could an expert retrain this tool to meet changing needs (e.g., to adhere to a new policy or to integrate new information)?
- Does the vendor enable operational experimentation and iterative phases of work?

If the operational experimentation is successful, with the success criteria met, and end-users deeming the system to be effective, then procurement can choose whether or not to consider other systems. The operational experimentation may be unsuccessful for a variety of reasons, such as the end-users not being able to complete their tasks on their own, the system not providing them with confidence that it was able to support their needs, or the system seemed untrustworthy. In these cases, the acquisition team should eliminate the system from consideration and move on to the next solution. This process provides quick and relevant feedback to the procurement team and reduces the chances of wasting funds on the wrong AI system.

Continuity of Capabilities

Implementation of a new AI system is just the beginning. The capabilities must continue to be available to the workforce for the system to be successfully adopted and integrated into existing processes. Systems are typically rejected when the interface design and interactions diverge from the end-users needs. Connecting with end-users ensures that aspects of design such as trustworthiness and transparency are interpreted and implemented appropriately. When end-users understand the systems' capabilities and limitations and are confident using it in context, it is likely to be successful.

The workforce will need to define processes and responsibilities for the following aspects of the system (Gardner et al., 2023):

- Continuous monitoring, test, evaluation, verification and validation practices such as Derr et al. (2025) and NIST (2023)
- Continuous performance monitoring appropriate for developers and end-users
- Risk mitigation planning and implementation
- Operations for training, fine-tuning, auditing, etc. of data and models such as DeCapria (2024)

As the AI system is adopted, the workforce may need to develop guidance for productive system use and even specify the systems' recommended uses and limitations. The user and development needs will change over time, so keeping an agile mindset will help the team to respond as needed.

Conclusion

This streamlined practice to inform procurement teams, enable operational experimentation, and ensure the continuity of capabilities will enable the Navy to more quickly implement effective AI systems that connect and augment human abilities. Better AI system procurement practices will enable the United States to solidify its position as the leader in AI and secure a brighter future for all Americans.

References

Armstrong, K. (2023, May 27). *ChatGPT: US lawyer admits using AI for case research.* BBC. https://www.bbc.com/news/world-us-canada-65735769

Carleton, A., Ivers, J., Ozkaya, I., Robert, J., Schmidt, D., & Zhang, S. (2025, February 27). *Perspectives on generative AI in software engineering and acquisition*. https://doi.org/10.58012/gtk0-b084

Dastin, J. (2018, October 10). *Insight - Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/

DeCapria, D. (2024, November 4). *Introduction to MLOps: Bridging machine learning and operations*. https://doi.org/10.58012/zkgg-en08

Defense Acquisition University. (2020). *Software acquisition*. DAU Adaptive Acquisition Framework https://aaf.dau.edu/aaf/software/

Defense Information Systems Agency. (2025). *DISA data strategy FY25-27*. https:///DISA-FY25-27-Data-Strategy

Derr, A., Echeverría, S., Maffey , K. R., & Lewis, G. (2025, February 17). *Introducing MLTE: A systems approach to machine learning test and evaluation*. https://doi.org/10.58012/cqzv-wj37

DrivenData. (n.d.). *Deon ethics checklist*. Retrieved June 20, 2024, from https://deon.drivendata.org/

Gardner, C., Robinson, K.-M., Smith, C. J., & Steiner, A. (2023, July 17). *Contextualizing end-user needs: How to measure the trustworthiness of an AI system*. Carnegie Mellon University, Software Engineering Institute. https://doi.org/10.58012/8b0v-mq84

Gartner. (2024). *Gartner survey finds generative AI is now the most frequently deployed*. https://www.gartner.com/en/newsroom/press-releases/2024-05-07-gartner-survey-finds-generative-ai-is-now-the-most-frequently-deployed-ai-solution-in-organizations

Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H., & Crawford, K. (2021, December 1). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92. doi:10.1145/3458723

Hegseth, P. (2025, March 10). *Modern software acquisition to speed delivery, boost warfighter lethality.* DoD. https://www.defense.gov/News/News-Stories/Article/Article/4114775/modern-software-acquisition-to-speed-delivery-boost-warfighter-lethality/

Martelaro, N., & Ju, W. (2020). What could go wrong? Exploring the downsides of autonomous vehicles. *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '20).* Association for Computing Machinery. doi:https://doi.org/10.1145/3409251.3411734

McDonald, K. (2023). *What is a minimum viable product (MVP)?* Agile Alliance. https://www.agilealliance.org/glossary/mvp/

Messeri, L., & Crockett, M. (2024, March 6). Artificial intelligence and illusions of understanding in scientific research. *Nature, 627*, 49–58. https://doi.org/10.1038/s41586-024-07146-0

Moran, K. (2024). *Usability (user) testing 101*. Nielsen Norman Group. https://www.nngroup.com/articles/usability-testing-101/

NIST. (2023). *NIST AI RMF: Artificial intelligence risk management framework (AI RMF 1.0).* https://doi.org/10.6028/NIST.AI.100-1

Smiley, L. (2022, March 8). *'I'm the operator': The aftermath of a self-driving tragedy*. Wired. https://www.wired.com/story/uber-self-driving-car-fatal-crash/

Tao, C. (2024, August). *Do not use LLM or generative AI for these use cases*. Towards AI. https://pub.towardsai.net/do-not-use-llm-or-generative-ai-for-these-use-cases-a819ae2d9779

U.S. Environmental Protection Agency. (2025, February 13). *Exploratory data analysis*. https://www.epa.gov/caddis/exploratory-data-analysis

Digital Engineering, Understanding the Policy and the Engineering in the Minimal Viable Product Approach

N. Peter Whitehead, PhD, ME—is a Senior Engineer and Professor of Policy Analysis at RAND, where he conducts research in systems engineering applications & practice, systems analysis, artificial intelligence, cybersecurity, supply chain risk management, and space technology. Prior to coming to RAND, Whitehead conceived of and led an AI at the edge research program at MITRE and supported the DHS HART biometric database program. He was an AAAS S&T Policy Fellow at NSF and OSTP, self-funded his PhD program in systems at UVa, served as Program Manager of an Air Force satellite program at Lockheed Martin, was Test Director of the Air Force DMSP mission sensor, and was a Foreign Service Officer, serving in 60 countries. Whitehead's PhD dissertation in systems developed an AI-supported empirical approach to determining the quality of a systems approach. [nwhitehead@rand.org]

Abstract

This analysis endeavors to clarify the dichotomy of policy and engineering in DoD system acquisition. It considers the Software Acquisition Pathway (DoDI 5000.87) in current DoD policy, approaching that policy from the perspective of good systems engineering practice. It endeavors to provide a bit of guidance on the following: distinguishing policy from engineering – using DoDI 5000.97 Digital Engineering as an example, distinguishing engineering writ-large from software coding, and understanding the importance of working closely with the stakeholder through the minimal viable product (MVP) process. It defines through allusion two distinct flavors (definitions) of MVP – the flavor practiced in commercial industry by many large software companies (systems engineering goal development), and the flavor directed specifically by DoDI 5000.87 (policy). This analysis attempts to show how to use them respectively in the acquisition policy flow and in the systems engineering process.

Summing-up: In developing capability for the DoD, there is a right way, a wrong way, and a policy way, and an acquisition program has to always understand which is in play.

Digital Engineering

The origins of the DoD digital engineering paradigm trace their lineage to the structured software architectures of the 1970s pioneered by such developers as Tom DeMarco and Edward Yourdon.[1] These approaches spawned the concept of computer-aided software engineering tools in the days before recent advancements such as Curser and OpenAI Codex (O'Regan, 2013). In parallel, the paradigm of object-oriented software development evolved to where, in 1995, Grady Booch, Ivar Jacobson, and James Rumbaugh integrated multiple conventions of software engineering and architecture into UML. Their goal was to construct an object-based programming tool where lines of code are replaced with objects, thereby simplifying and expediting the coding process. As it turned out, UML required an extreme level of detail and effort that paradoxically made line-by-line coding more efficient, and so it failed to be adopted for its intended purpose (Bell, 2004; Pandey, 2010). The front-end structured approach to software development was also overshadowed by the Agile approach at the dawn of the new millennium, making architectural frameworks a tool for later documentation but not useful for the new approaches to development (Whitehead et al., 2024).

[1] Systems engineering and the software-based paradigm that became digital engineering diverged primarily through the work of Barry Boehm at USC, the inventor of the systems engineering Vee diagram. Boehm effectively postulated and promoted the assumption that all systems behave like software code, so coding-based systems analyses would be close enough.
The Vee started as a greater than symbol and is wholly derived from software development practices in the 1970s, not established systems engineering practice. See Boehm (1981, 1984).

MODAF and DoDAF

Meanwhile, in Britain, engineers in the Ministry of Defense developed a graphical approach to describing complex systems called the Ministry of Defence Architectural Framework. By the year 2000, this had become the U.S. standard known as DoDAF, and the software tool System Architect was adopted as the industry standard for creating the multi-layered weapon system program perspectives of DoDAF (DoD Chief Information Officer, 2021).

In the early-mid 2000s, a group of software architects, seeing the similarity between DoDAF and the graphical products of UML, created a dialect of UML that they called Systems Modeling Language, thus creating an open-source alternative to System Architect (SysML.org, n.d.). SysML was adopted by a software-centric engineering organization, INCOSE, which coined the term *MBSE* to describe the DoDAF-like architecture use of SysML.[2] In 2006, the Office of the Under Secretary of Defense for Acquisition and Technology, MITRE, Lockheed Martin, Boeing, and others collectively aligned on the MBSE initiative as proposed by INCOSE for system architecture applications in weapon system programs (Hardy, 2006).[3]

Model-Based

The term *model-based systems engineering* might catch some experienced systems engineers off guard, as all systems engineering through thousands of years of practice has been model based, making the term itself sound redundant.[4] Egyptians built scale models of pyramids to study the related mathematics, engineer their construction, and plan the required logistics (Rossi, 2004). Galileo developed mathematical models of the parabolic trajectory of cannon shells that proved to be highly accurate in practice (Naylor, 1976). Bell Labs practiced what Arthur D. Hall (1962) called *systems engineering* and defined it as “organized creative technology and its functions” (p. 3). NASA and military engineers and program managers leveraged thousands of models in successfully putting men on the moon and giving rise to the current perceived value of good systems engineering practice in a complex program (Miles, 1974).

---

[2] MBSE is “the formalized application of modeling to support system requirements, design, analysis, verification and validation activities beginning in the conceptual design phase and continuing throughout development and later life cycle phases” (INCOSE, 2007, p. 15). See also SysML.org (n.d.).

[3] Hardy (2006) writes:

MBSE enhances the ability to capture, analyze, share, and manage the information associated with the complete specification of a product, resulting in the following benefits:

- Improved communications among the development stakeholders (e.g. the customer, program management, systems engineers, hardware and software developers, testers, and specialty engineering disciplines).
- Increased ability to manage system complexity by enabling a system model to be viewed from multiple perspectives, and to analyze the impact of changes.
- Improved product quality by providing an unambiguous and precise model of the system that can be evaluated for consistency, correctness, and completeness.
- Enhanced knowledge capture and reuse of the information by capturing information in more standardized ways and leveraging built in abstraction mechanisms inherent in model driven approaches. This inturn [sic] can result in reduced cycle time and lower maintenance costs to modify the design.

[4] Arthur D. Hall describes the origins of the concept that became labeled *systems engineering* at Bell labs in describing the 1940 development of the TD-2 radio relay system: “the name was new, but the functions were not.” He traces *systems analysis*, which includes what we will call system goal definition later in this report, to a philosophy developed in the 1940s by the RAND Corporation. Hall adds the terms *systems thinking* and *systems approach* to the list of supporting concepts. These concepts had existed and evolved over millennia, so, while the authors use the term *systems engineering* here, it is used in a general sense to include the supporting concepts and the history of systems engineering predating 1940 (Hall, 1962, pp. 7, 26).

We know from our study of the state of practice in the DoD that the term *MBSE* describes the leveraging of object-oriented architecture modeling, specifically SysML, as derived from the waterfall, object-oriented software development practice of the late 1990s (Hardy, 2006). In other words, it is a 1990s-era software coding tool repurposed through policy to serve as an engineering tool.[5] We have observed SysML used for architecture, system interface, and organizational modeling predominantly, with teams leveraging the tools for other applications as they see fit. We also know that stakeholders across the DoD and the defense industrial base define the details and scope of MBSE differently, so no observation on our part may be considered universal. Further confusion of systems versus software emanates from the IEEE Software Society standards (e.g., ISO/IEEE 15288 and 24641) that are adopted incorrectly by some in the DoD and in the defense industrial complex as systems engineering approaches (Whitehead, 2024).

Studies conducted by a DoD-sponsored university affiliated research center, SERC, in the 2010s worked to leverage MBSE as defined in SysML onto metamodel optimization concepts originated by such researchers as Markish and Willcox (2003) and Kühne (2006). This work led ultimately to the concept of digital engineering as espoused in the 2018 policy document *DoD Digital Engineering Strategy* and DoDI 5000.97*, Digital Engineering (2023)* (Bone et al., 2019; Shyu, 2023).

Distinguishing Policy from Engineering

Based on the above, digital engineering and model-based systems engineering as prescribed in DoDI 5000.97 are policies – they come to us from the top-down, and their use is not supported with empirical evidence. Policies have terms of justification reflecting the lack of evidence, such as "***can*** modernize how the DoD designs, develops, [etc.]" and "***should*** enable faster, higher-quality decision making" (Shyu, 2023, p. 8).

Engineering emanates from the bottom up, driven by design (goals) and applied science – empirical evidence. Momentum *equals* mass times velocity. Increasing pressure *will* reduce the volume of a gas if the temperature is constant.

Confusion may emanate from the tendency of some documents such as DoDI 5000.97 to readily conflate the two, as in this statement:

> *Digital engineering requires planning and providing financial and other resources for digital methods (e.g., model-based systems engineering (MBSE), product life-cycle management, computer aided design) in support of program activities to the maximum extent possible.* (Shyu, 2023, p. 3)

In that statement, *digital engineering*, *MBSE* and *product life-cycle management* are policies; *computer-aided design* is an engineering tool, the use of which is supported with empirical evidence.

Goal Definition in Concept Development – The Engineering

Engineering starts with system goal definition, a highly complex, human-centric endeavor with no direct parallel in digital engineering policy. Goal definition also brings into play the stakeholder interaction known as minimal viable product (MVP).

[5] Coding and engineering writ-large are two very different practices. The computer does as instructed. Code may be complicated, but it does not deal with complexity in the systems engineering sense. Complexity, as dealt with in systems engineering, involves humans, politics, the axiological as well as the applied science and the interaction of often incompatible sub systems. This will be addressed further in the goal development section.

Before initiating the planning phase, the goal definition phase of the program can be the most important phase of any development program (Gibson et al., 2016). This phase examines where the program must functionally go, how to measure progress, who will benefit, and how they will benefit. It provides the foundation for planning the way the system will operate, the path, and the methodologies (Buede & Miller, 2016). In software acquisition, goal definition does not end at any milestone but continues iteratively through the entire system lifecycle. The software is never finished, and goals evolve. Getting the preliminary step of goal definition correct will reduce program risks and streamline both complex and rudimentary aspects of the acquisition process (Whitehead, 2014).

Goal definition defines where to go and how we can tell when we arrive. Engineering necessarily follows with how to get there. Policy puts necessary constraints on engineering.

Identifying Stakeholders

The originating office must not define the system goals in a vacuum, no matter how well they may understand the problem. At the outset of goal definition, they should establish a hierarchical list that includes all of the system stakeholders (Gibson et al., 2016, pp. 55–75). At the top of the hierarchy are the end users, the customers. Next are the entities that support the end-users directly, to include their help desk functions, financial representatives, and the many, varied sources of data and models for their simulations. Next are the enterprise entities that will be responsible for training, cloud operations, future planning, access, security, and integration across the military enterprise. The originating or coordinating office is not necessarily the provider of the facilities or resources needed to make the program go but is the critical center of this and most other following activities.

Elucidating Stakeholder Goals via Scenario Development

In facilitated exercises conducted multiple times, the coordinating office aligns and integrates the goals of all the respective stakeholders into a common set of functional goals.[6] These goals take the form of multiple scenarios describing the use, function, and implementation of the system (Alexander & Maiden, 2005). Systems engineering shows us a litany of approaches for doing this, generally parsed into preliminary surveys, in-person exercises, hotwash, iteration, and final drafts socialized for stakeholder comment.[7] An important principle the coordinating office should adhere to is to focus the process of scenario development on a clear articulation of the desired end state, rather than identifying specific parts of the design and/or specific development approaches (Weinberg, 1982). Instead, the development team must continually refine the development features and process in the (later) planning and execution phases in the context of the stakeholder-informed scenarios (Reis, 2011, pp. 99–113). Figure 1 suggests how a classical system development process—one that does not specify exactly how the desired system will be used and by whom—can ultimately miss the desired end state, the conflict of policy and engineering.

[6] An objective, outside entity could provide neutral facilitation based on systems principles. These functional goals are likely to differ from the conceptual system architecture shown in Figure 1.

[7] Instead of seeking approval from all stakeholders, the coordinating office will likely adjudicate comments using a formal process such as Department of Army Form 7874, the Army-Wide Staffing Comment Resolution Matrix.

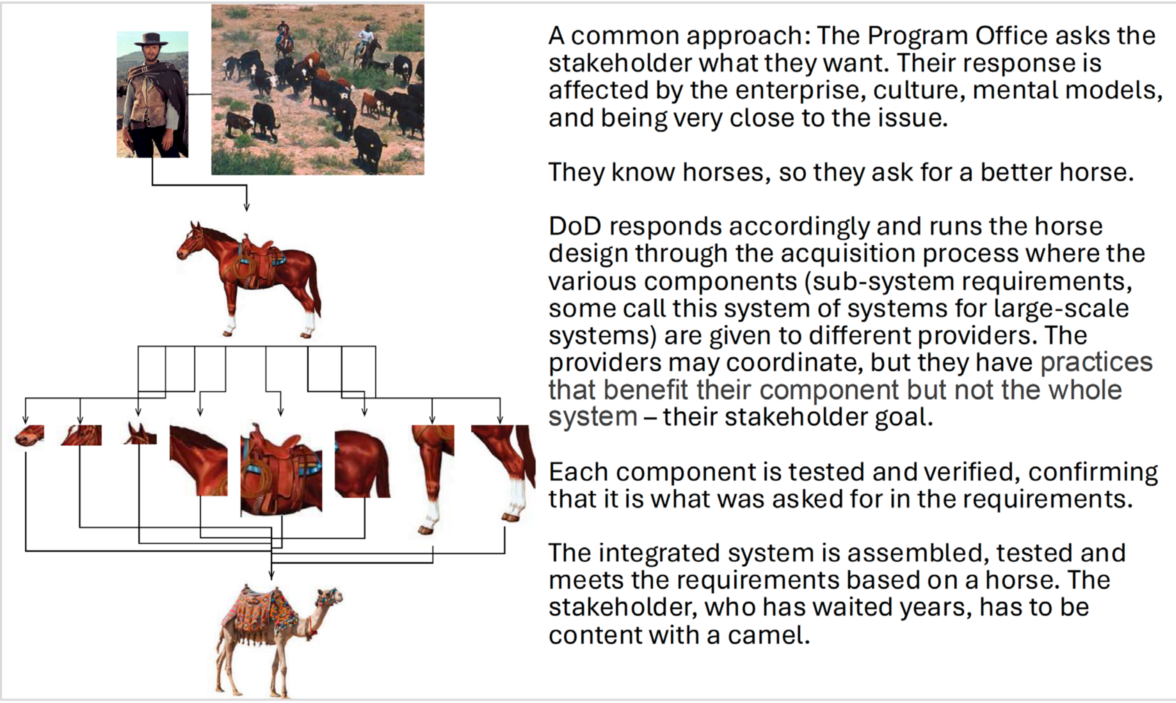


Figure 1. Classical System Development Failure

Scenarios generally have four parts described in plain language or with relatable examples. First, describe the environment where the system will be used, including accessibility hardware and the operational conditions (e.g., in a remote location on a tablet with limited, unclassified connectivity or in a training center in the continental United States with high-speed, classified connectivity).

Second, the users of the simulation are characterized to scope their familiarity with the intended use. This can be achieved with various traits but should be definable and explicit, such as domain-specific training, service experience, and typical operating tempo, among many others.

Third, the immediate user-goals and intended activities are described (e.g., test the simulated effectiveness of a new counter unmanned aerial vehicle system or improve the logistics and timeline of a deployment of armored personnel carriers to Europe).

Finally, the scenario should specify concrete outcomes and/or data expected from the simulation activity, such as gaining skill training, refining a conceptual design, developing or validating novel concepts of operation, planning an upcoming operation, or assessing weapon effectiveness (Alexander & Maiden, 2005).

The set of goal definition scenarios is intended to be as complete as possible, in recognition of the iterative minimum viable product (MVP) process, illustrated in Figure 2.[8] The MVP process in industry is characterized by the phrase, *not like that, more like this,* as options are presented to the stakeholders by the developers (Reis, 2011, pp. 99–113). Each successive hypothetical design is less wrong, iteratively both refining and explaining the mental model of

[8] MVP is often interpreted differently in different circles. The first part of this chapter loosely follows the industry definition of an MVP, one that is developed closer to the goal definition phase and provides the simplest product viable for commercialization. The second part of this chapter focuses on the DoD definition, where iterative products are delivered during the execution phase.

the stakeholder to the developer in terms that they both understand, as in Figure 2.[9] Not only does the creation of end-user scenarios help to refine the initial design, but it establishes stakeholder buy-in for the program, not unimportant in DoD culture. Assumptions, beliefs, and unwritten cultural factors will inevitably have an impact on the development. These axiological factors represent a critical subject for discussion in the goal definition process that should not be ignored.

Note that this is the *first* of the two flavors of the MVP process to be employed during the system development and acquisition process flow. This is the system goal definition phase before the DoDI 5000.87 Capability Needs Statement/planning phase. Subsequently, the DoD flavor of MVP will be employed during the DoDI 5000.87 execution phase.

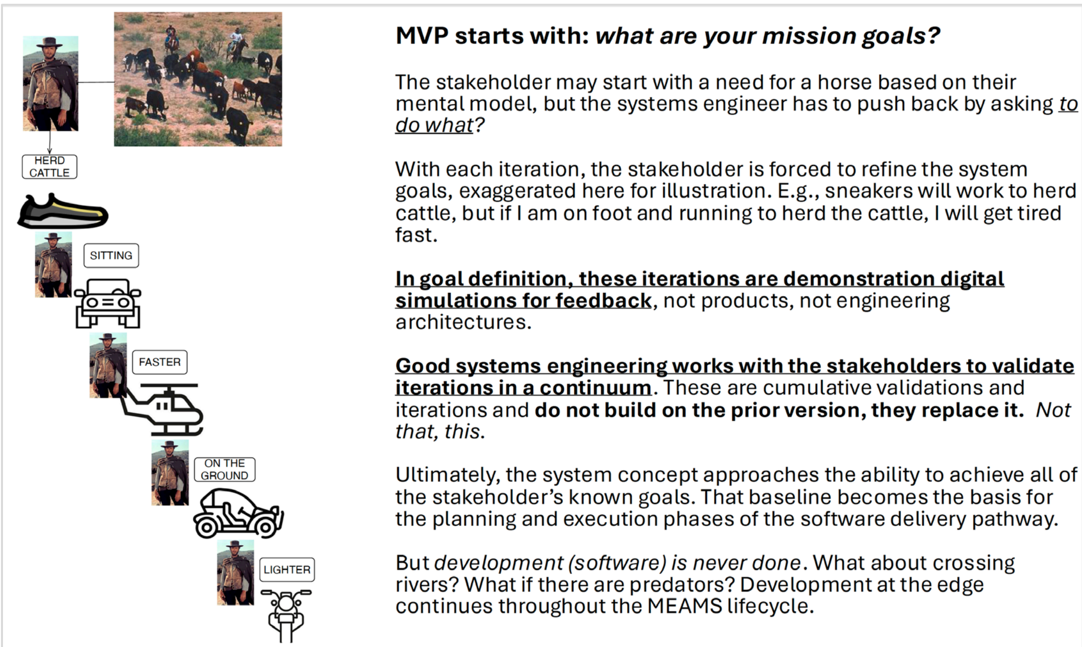


Figure 2. Minimum Viable Product System Development – Commercial Industry Flavor

System Scope

The system scope is expanded by the developed scenarios but bounded by the management triad of schedule, cost, and quality. How much can be afforded, when do stakeholders need it, what are the minimum quality attributes that get the job done for them, and what are acceptable program risks? In the goal definition phase, the coordinating office accumulates these data for planning and costing, including all the logistics and support for deployment and lifecycle operation. Financial bounds have to be a part of the goal definition process to obtain an on schedule and within cost (i.e., viable) end state.

Indices of Performance

Indices of performance (IoPs) represent metrics that relate to the respective goals. These metrics can be technical, descriptive of a process, or concerning the engineering or execution activities themselves. IoPs will be subject to refinement during the MVP process, but the key, longitudinal IoPs should be maintained throughout the lifecycle, and the units must be consistent. Critical aspects of IoPs are that they be measurable, objective, nonrelativistic, meaningful, and understandable to the stakeholders (Gibson et al., 2016, pp. 41–45).

[9] Systems engineers have known this approach for a very long time; see Churchman (1968).

Software Acquisition Pathway – The Policy

Once system goals are well defined, the program advances in the DoD acquisition cycle. DoDI 5000.02, *Operation of the Adaptive Acquisition Framework*, lays out the different pathways that can be used to acquire solutions for end users throughout the DoD (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2022). Figure 1 of DoDI 5000.02 includes software acquisition which is described in detail in DoDI 5000.87 and reproduced in Figure 3.[10] Compared with the other acquisition pathways, which are largely unidirectional and marked by milestone events, software acquisition is iterative, implying that software components must be continuously improved (via MVP iteration and CI/CD) during the entire system lifecycle.

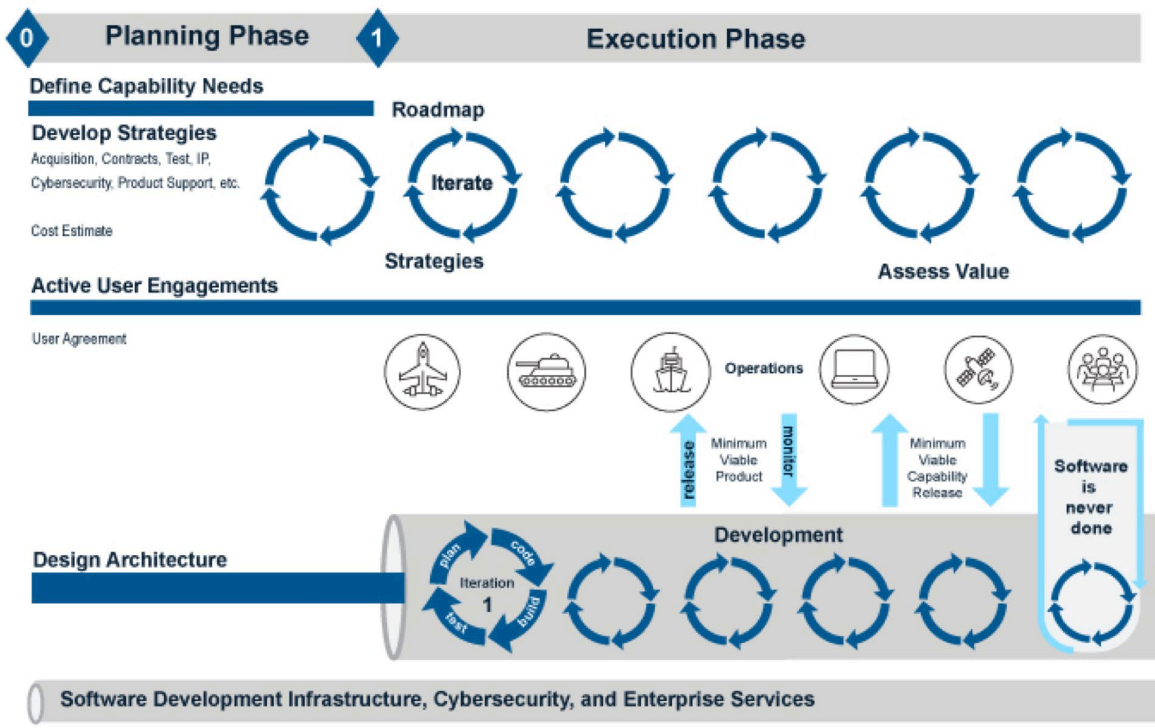


Figure 3. The Software Acquisition Pathway
(Lord, 2020, p. 8)

As shown in Figure 3, the software acquisition pathway is divided into two phases, planning and execution. During the initial planning phase, market analysis is conducted to determine if a commercial off-the-shelf (COTS) software solution to address the system goals is available for purchase. If available, the COTS option must be pursued in accordance with Title 10, Section 3453. If COTS is not available, then the planning phase of the framework proceeds to understand end user needs and establish methodologies to deliver to the users the correct software capabilities (Lord, 2020, p. 9).

Programs using the software acquisition pathway will be identified in competent DoD program lists and databases within 60 calendar days of initiating the planning phase in accordance with the DoD's implementation of Section 913 of Public Law 115-91 on acquisition data analysis (Lord, 2020, p. 10).

[10] The actual process requirements will largely be defined by the overall cost, application, and/or ownership of the final system. This, in turn, will specify the funding source, proponent, and coordinating office.

The Planning Phase

The planning phase of the Software Acquisition Pathway is guided by a draft Capabilities Need Statement (CNS) that is developed by the operational community via the MVP process described above. Through the process, requirements in the CNS are re-prioritized to facilitate effective software development, and user engagement is utilized to update the CNS accordingly. The decision authority, the Under Secretary of Defense for Acquisition and Sustainment, selects the project manager to strategize and govern the software acquisition process (Lord, 2020, p. 9). Software design and architecture attempt to use existing enterprise services as much as reasonably possible. However, this should be guarded by focusing on the system goals discussed earlier in the chapter; planners should realize when the re-use benefit of existing solutions is outweighed by the gaps to goals introduced when forcing alignment. Planning considers and documents in appropriate artifacts a range of factors including but not limited to development environment, automation tools and capabilities, cybersecurity threats, risk-based lifecycle management, testing, and evaluation (Lord, 2020, p. 10). Once the decision authority validates that the appropriate acquisition artifacts are complete and the strategies, analysis, and resources are in place, the process transitions to the execution phase. From planning and through execution, the program develops and tracks metrics of success and keeps cost estimates, costs, and software data reporting up to date.

Other Required Planning Documents

In conjunction with the CNS, a user agreement, acquisition strategy, intellectual property strategy, test strategy, and cost estimate must be approved to transition to the execution phase. The sponsor and program manager must develop a user agreement to ensure commitment, involvement among parties, and delegate decision-making authorities (Lord, 2020, p. 11). Decisions include capabilities defining, capabilities prioritization, software feature trade-offs, software cadence, user acceptances, and readiness for operation deployment. In addition, the user agreement will commit proper resourcing to engage users and create a system for feedback and ways to shape requirement details.

The Acquisition Strategy

The acquisition strategy identifies how to acquire, develop, deliver, and sustain software capabilities for the end users' needs (Lord, 2020, p. 11). Active collaboration between the program manager, program stakeholders, and functional experts ensure the acquisition strategy addresses current environments, priorities, risks, and approaches. The acquisition strategy will be revised by the program manager until it is sufficient for the decision authority to approve development and continue to mature it through the acquisition lifecycle. The acquisition strategy will be approved by the decision authority to include process and documentation tailoring. Key elements of the acquisition strategy are risk-based business and technical management, roadmap and cadence for delivery, flexible and modular contract strategy, planned government personnel and resources, tailoring to use modern practices, high-level test strategies, architecture strategy to enable open modular systems, intellectual property training, product support strategies, and program manager strategy to ensure all is in accordance with law and regulations. If software is embedded, then it must align with the platform acquisition strategy.

The Intellectual Property Strategy

The intellectual property strategy (IPS) identifies and describes the management of delivery and license rights for all software and related material necessary to meet requirements (Lord, 2020, p. 12). The IPS must support and be consistent with government strategies and implemented through requirements in contracts. Rights and obligations of the government and industry should be understood by the program manager to handle strategy and negotiation for software deliverables and license rights. The IPS includes negotiation for and periodic delivery of software components (Lord, 2020, p. 13). The IPS should address collaboration with other developers and users of software, in the case of government will take delivery and/or modify source code, to reduce duplication. The program manager should attempt to avoid the creation of program-specific versions of software components. Commercial or proprietary software not previously included in the IPS will be approved by the program manager before insertion into software developed for the government. The IPS identifies where intellectual property may result from government investments and treat them appropriately. The program manager should require delivery of all source code at the government's expense and any other requisite documentation. Timelines for delivery should be planned around transitions to new contractors or the government.

Test Strategy

The test strategy defines the process by which capabilities, features, user stories, use cases, and elements will be tested and evaluated to demonstrate if criteria are satisfied. The test strategy identifies the independent test organizations, testing artifacts that will be shared, tools and resources for data collection, and transparency. Tests should assess software performance, reliability, sustainability, and other key metrics. For embedded software, safety assessments and mitigation strategies should be included for any implications to the overarching system. The schedule for embedded software should also align with test and integration for the overarching system. To the extent practical, testing and operational monitoring should be automated for user evaluation. The test strategy should include information in accordance with applicable modeling and simulation policies. The decision authority will approve the test strategy, and the Director, Operational Test and Evaluation, will be the final approver for programs on their oversight list.

Cost Estimate

The cost estimate, in accordance with DoDI 5000.73, *Cost Analysis Guidelines and Procedures*, estimates and considers the technical content of the program described by the other software acquisition pathway required documents. The initial cost estimate must be completed before the execution phase and then updated annually. Where applicable, cost and software data reporting, to include software resources data reports, must be submitted in accordance with DoDI 5000.73 policies and procedures.

The Execution Phase

Software capabilities that correspond with the needs of the end users are developed and delivered during the execution phase. The program assembles components from enterprise services and contracts. Existing connections are preferred to new ones and based on the acquisition and intellectual property strategies. The program maximizes automation of processes related to the project when possible. Consideration should be given for lifecycle objectives and managing technical debt. The sponsor and program office develop and maintain a product roadmap, while the product owner and office maintain a backlog detailing a prioritized list of user needs. The product roadmap and backlog are shaped by continual user feedback.

That continual user feedback takes place in the second iteration of the MVP process, this time during acquisition execution. The program manager and sponsor use an interactive human-centric design process to define the MVP as user needs evolve. If the MVP does not have sufficient capabilities or performance to deploy into operations, then the program manager and sponsor define an MVP release. The MVP release delivers initial capabilities to enhance mission outcomes and must be deployed to an operational environment within a year of the first obligated funds given to acquire or develop new capabilities. Subsequent MVP releases should be delivered at least annually per policy. Through execution, the program should continually update the development process and take user feedback to inform short-term capability deliveries and long-term solutions. Testing should be guided by risk strategies, and cyber testing and monitoring should be automated to be used to support a conditional authority to operate or accelerated accreditation process.

Cybersecurity policies and assigned authorized officials guide this process. Recurring cybersecurity assessments should be performed on all components of the process. Program managers work with stakeholders to provide controls to enable conditional authority to operate where needed and ensure secure development, cybersecurity and assurance capabilities, and secure lifecycle management. Intellectual property strategy considerations should be marinated through execution. Programs develop and track metrics of success of the program, keep cost estimates, costs, and software data reporting up to date from planning and through the execution phase. The sponsor and user community conduct value assessments at least on delivered software. The results of the assessments inform progress and updates to the process.

Summary/Recommendations

- Both policy and engineering impact acquisition programs. Please, never confuse the two.
- Understand the limits of software coding tools such as UML and SysML and don't conflate the policy directive to use them in programs with systems engineering practice – despite the confusing verbiage.
- The DoD Software Acquisition Pathway includes plenty of space to accommodate a wide variety of sound systems engineering approaches to deliver capability to the warfighter. We have attempted to show one systems engineering approach to doing so via MVP.
- Program success will likely depend on the regular engagement of the stakeholders in the development and iteration process via MVP and the budgeting of continuous improvement over the lifecycle of the program.

Bibliography

Alexander, I. F., & Maiden, N. (2005). *Scenarios, stories, use cases: Through the systems development life-cycle*. John Wiley & Sons.

Bell, A. E. (2004). Death by UML fever: Self-diagnosis and early treatment are crucial in the fight against UML fever. *Queue*, *2*(1), 72–80.

Boehm, B. W. (1981). Software engineering economics. *New York*, *197*.

Boehm, B. W. (1984). Software engineering economics. *IEEE Transactions on Software Engineering,* (1), 4–21.

Bone, M. A., Blackburn, M. R., Rhodes, D. H., Cohen, D. N., & Guerrero, J. A. (2019). Transforming systems engineering through digital engineering. *Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, *16*(4), 339–355. https://doi.org/10.1177/1548512917751873

Buede, D. M., & Miller, W. D. (2016). *The engineering design of systems: Models and methods* (3rd ed.).

Churchman, C. W. (1968). *The systems approach* (Vol. 1). Dell.

Defense, D. C. I. O. U. S. D. o. (2023). *DODAF - DOD architecture framework version 2.02.* https://dodcio.defense.gov/library/dod-architecture-framework/

DoD. (2020). *Operation of the adaptive acquisition framework*.

DoD Chief Information Officer. (2021). *The DODAF architecture framework version 2.02*.

Friedenthal, S., Griego, R., & Sampson, M. (2007). INCOSE model based systems engineering (MBSE) initiative. *INCOSE 2007 symposium*.

Gibson, J. E., Scherer, W. T., Gibson, W. F., & Smith, M. C. (2016). *How to do systems analysis: Primer and casebook*. John Wiley & Sons.

Hall, A. D. (1962). *A methodology for systems engineering*. D. Van Nostrand Company.

Hardy, D. (2006). Model based systems engineering and how it aids DoD acquisition & systems engineering. *9th Annual Systems Engineering Conference*.

INCOSE. (2007, September). *Systems engineering vision 2020.*

Kühne, T. (2006). Matters of (meta-) modeling. *Software & Systems Modeling*, *5*(4), 369–385.

Light, T., Younossi, O., Clayton, B., Whitehead, P., Wong, J. P., Pfeifer, S., & Triezenberg, B. L. (2022). *A preliminary assessment of digital engineering implications on weapon system costs*. T. R. Corporation.

Lord, E. M. (2020). *Operation of the software acquisition pathway* (DoD Instruction 5000.87). https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500087p.PDF

Markish, J., & Willcox, K. (2003, October). Value-based multidisciplinary techniques for commercial aircraft system design. *American Institute of Aeronautics and Astronautics Journal*, *41*(10), 2004–2012. https://doi.org/10.2514/2.1890

McQuade, J. M., Murray, R. M., Louie, G., Medin, M., Pahlka, J., & Stephens, T. (2019). *Software is never done: Refactoring the acquisition code for competitive advantage, final report of the software acquisition and practices (SWAP) study conducted by the Defense Innovation Board*. https://media.defense.gov/2019/Apr/30/2002124828/-1/-1/0/SOFTWAREISNEVERDONE_REFACTORINGTHEACQUISITIONCODEFORCOMPETITIVEADVANTAGE_FINAL.SWAP.REPORT.PDF

Miles, R. F., Jr. (1974, January 15). *A contemporary view of systems engineering*. Jet Propulsion Laboratory.

Naylor, R. H. (1976). Galileo: The search for the parabolic trajectory. *Annals of Science*, *33*(2).

Office of the Under Secretary of Defense for Acquisition and Sustainment. (2022, June 8). *Operation of the adaptive acquisition framework* (DoD Instruction 5000.02, change 1).

O'Regan, G. (2013). Ed Yourdon. In *Giants of computing: A compendium of select, pivotal pioneers* (pp. 277–279). Springer.

Pandey, R. (2010). Architectural description languages (ADLs) vs UML: A review. *ACM SIGSOFT Software Engineering Notes*, *35*(3), 1–5.

Reis, E. (2011). The lean startup. *Crown Business*, *27*, 2016–2020.

Rossi, C. (2004). *Architecture and mathematics in ancient Egypt*. Cambridge University Press.

Shyu, H. (2023). *Digital engineering* (DoD Instruction 5000.97). DoD. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500097p.PDF?ver=bePIqKXaLUTK_Iu5iTNREw%3D%3D

SysML.org. (n.d.). *SysML open source project: What is SysML? Who created it?*

Weinberg, G. M. (1982). *Rethinking systems analysis and design*. Little, Brown.

Whitehead, N. P. (2014). *The dimensions of systems thinking - An approach for a standard language of systems thinking* [Doctoral dissertation, University of Virginia]. https://libraetd.lib.virginia.edu/public_view/3x816m893

Whitehead, N. P., Light, T., Luna, A., & Mignano, J. (2024). *A framework for assessing the costs and benefits of digital engineering*.

Employing a Variable, Portfolio Contract Model to Accelerate Innovation Incorporation, Enhance Operational Sustainability, and Reduce Supply Risk in the Procurement Process

Lieutenant Commander Adam Bradley Pierce—LCDR Pierce is a USN Explosive Ordnance Disposal Officer, currently stationed at the Navy Expeditionary Warfighting Development Center, in Virginia Beach, VA. He is a graduate of the Naval Postgraduate School (DDM-Financial Management, 2022). [adam.b.pierce.mil@us.navy.mil]

Abstract

This paper focuses on providing solutions to two problems plaguing federal acquisition processes: (1) limited ability to rapidly incorporate innovation into contracted procurement and (2) supply risk associated with high-volume quantity contracts that engage individual contractors. The author proposes solutions which are designed to resolve these problems, which include (a) capability requirement documents written to increase design flexibility and uniformity in operations and maintenance, and (b) a variable, portfolio contract model which simultaneously engages multiple contractors—to increase overall contract production capacity and reduce supply risk—and is able to change the quantity demanded from each contractor ("market share"), based on innovative improvements to cost, schedule, and/or performance. The contract model is applied to a high-end, near-peer, maritime competitive environment, requiring high-volume procurement. An evaluation of the contract model, consisting of 120 individual simulations, demonstrated (i) increased contract production capacity, (ii) consistent increase in procurement quantity demanded from innovative contractors, and (iii) increased product performance rating—which translates to a higher quality capability delivered to Warfighters. The paper concludes with a recommendation to implement a variable, portfolio contract to ensure timely, risk-mitigated delivery of high-volume, high attrition capability for the future, maritime fight.

[T]he security environment is rapidly evolving, and the current PPBE process is not capable of responding as quickly and effectively as needed to support today's warfighter. The Department of Defense (DoD) needs a new process, one that enables strategy to drive resource allocation in a more rigorous, joint, and analytically informed way. The new process should also embrace changes that enable the DoD to respond effectively to emerging threats while leveraging technological advances. (Commission on PPBE Reform, 2024)

Background and Problem Statement

Background

Historically, defense acquisition has struggled with the adoption of and incorporation of innovative technologies—focused on improving the cost, schedule, and performance of acquisition programs—into materiel capabilities. This is due, in part, to the reliable but rigid nature of the Big "A" process.

Efforts to inject agility and adaptability into the process resulted in the development of the Adaptive Acquisition Framework (AAF), which provides multiple, tailored avenues for acquisition.

Despite the flexibility the AAF provides, one barrier to innovation remains. Once a contract is awarded, there is no incentive for contractors to innovate. This occurs because contractors develop and manufacture based on the Government's requirements. Overly prescriptive requirements documents define capabilities in a way that forces contractors to build systems that "are exactly this thing," as opposed to "a thing that is capable of accomplishing minimum operating requirements." Further, because the contract exists exclusively between the contractor and the Government, the element of competition—a key driver of innovation—is eliminated.

For certain acquisition programs—like aircraft carriers, submarines, and advanced aircraft—intra-contract competition may not be a factor. These programs produce relatively low numbers of systems, procured at higher costs, over long periods of time. However, for higher volume acquisition programs—like munitions or unmanned sensor platforms supporting hybrid fleet, maritime domain awareness—the ability to quantify the value of innovation, incorporate the innovation into a contracted capability, and reward a contractor for their investment would be invaluable in the quest to get the best equipment into the Warfighters' hands.

Another aspect of this discussion is the management of risk. Specifically, the focus is on supply risk and its impact to mission risk. In his 2003 *Journal of Purchasing and Supply Management* article, George Zsidisin defines supply risk as the "probability of an incident associated with inbound supply from individual supplier failures or the supply market occurring, in which its outcomes result in the inability of the purchasing firm to meet customer demand or cause threats to customer life and safety," (Zsidisin, 2003).

Supply chain disruptions observed during the COVID-19 pandemic demonstrated the vulnerability of production processes. In order for the Government to appropriately manage mission risk through supply risk mitigations—specifically for high volume acquisitions—it must employ contracts that can ensure the production capacity necessary to fulfill required operational capabilities <u>and</u> maintain surge capacity in reserve.

The center of gravity for injecting innovation and effectively managing risk is the relationship between the Government and contractors, which takes place in the form of a contract. As such, the Government must leverage the contract as a tool to incentivize innovation and reduce risk to acceptable levels.

Problem Statement

The following is a list of problems that drive the efforts of this research:

1. **Current conditions disincentivize innovation:** Under the current system, the Government bears the burden of creating the conditions that facilitate and enable innovation, including writing contracts that compensate contractors for their research and development expense. This runs contrary to private industry, in which competition forces you to innovate or die. The Government must employ contracting methods that place this burden squarely on the contractors and rewards the results of innovation—as opposed to compensating the effort.
2. **Overly prescriptive requirements inhibit innovation and adaptation:** The way requirements documents are written can either enable or hinder the ability for contractors to innovate during the term of the contract. Overly prescriptive requirements ("build this exact thing") inhibit innovation. The Government must write requirements in a way that establish minimum operating requirements <u>and</u> allow contractors the flexibility to "solve the problem" in their own way.
3. **Managing supply risk manages mission risk:** For acquisition programs that provide high volume materiel capability contracts, engaging with individual contractors increases supply risk. Since supply risk influences mission readiness, reductions to supply risk—via increasing the number of contractors engaged—translate into reductions to mission risk.

Scope and Goals

Scope

The scope of this research focuses on the development of contracting practices which incorporate competition as a driver of innovation, relies on capabilities requirements that do not inhibit innovation, and effectively reduce supply and subsequently mission risk.

Goals

To accomplish this, the following goals are established to drive research efforts:

1. **Research Goal 1:** Understand the current capabilities and limitations of Government contracting with regard to innovation generation
2. **Solution Goal 1:** Develop solutions which place the impetus of cost, schedule, and performance innovations in the hands of contractors
3. **Research Goal 2:** Understand the fundamentals of requirements documents and how these documents can either enable or hinder innovation
4. **Solution Goal 2:** Provide recommendations for writing requirements documents that enable innovation
5. **Research Goal 3:** Identify instances in which private industry has employed portfolio contracting and determine the resultant levels of success/failure
6. **Concept Goal 1:** Use knowledge gained from research to develop a contract model that achieves the aim of Solution Goal 1 and test the model under a range of scenarios
    a. **Contract model characteristics:**
        i. Able to engage multiple contractors (portfolio contract)
        ii. Able to rate individual contractor's proficiency in managing cost, schedule, and performance against all contractors engaged
        iii. Able to reward contractors who innovate by increasing the quantity of supply demanded (modifying the terms of the contract)—at the expense of the competing contractors
7. **Analysis Goal 1:** Analyze the results of model testing to assess the theoretical viability of the contract concept

Summary of Literature Review

The review of literature pertaining to the research goals was focused on three main areas: feasibility of employing a portfolio contract, requirements documents as adaptation-enablers, and identify instances in which private industry has employed portfolio contracts and the degree of success or failure experienced.

This section concludes with a discussion on concerns regarding the implementation of a portfolio contract.

Contracts

Federal Acquisition Regulations

The following summarizes research into the Federal Acquisition Regulations (FAR) focused on determining (1) if current regulations support a single contract engaging multiple contractors and (2) whether the mechanisms exist to alter the conditions of a contract during the term of engagement.

Indefinite-delivery contracts are contracts for supplies that do not procure or specify a firm quantity of supplies (other than a minimum or maximum quantity) and that provide for the issuance of orders for the delivery of supplies during the period of the contract. Essentially, they provide the Government with flexibility to increase/decrease demanded quantity supplied by the contract, based on changing operational conditions. This is beneficial for high volume acquisition programs (FAR 16.5, 2025).

Subordinate to indefinite-delivery contracts are two, applicable subgroups. The first, requirements contracts, provide for filling all actual purchase requirements of designated government activities for supplies or services during a specified contract period, with deliveries or performance to be scheduled by placing orders with the contractor. A critical caveat for requirements contracts states that "no requirements contract in an amount estimated to exceed $100 million (including all options) may be awarded to a single source unless a determination s executed in accordance with 16.504(c)(1)(ii)(D)," (FAR 16.503, 2025).

The second, indefinite-quantity contracts, provide for an indefinite quantity, within stated limits, or supplies or services during a fixed period. Quantity may be stated as number of units or as dollar values. Additionally, a subcomponent of indefinite-quantity contracts is the multiple award preference, which directs contracting officers to give preference to making multiple awards of indefinite-quantity contracts under single solicitation for the same or similar supplies to two or more sources, to the maximum extent possible (FAR 16.504, 2025).

Contract structure, as a limitation to building a portfolio contract, is addressed by the uniform contract format. Its core components: schedule, contract clauses, list of documents, and representations & instructions all provide the foundation on which to build a functional, portfolio contract (FAR 15.204, 2025).

The ability to alter the conditions of the contract—to reward an innovating contractor—exists in the form of contract modifications. Specifically, bilateral contract modifications provide the mechanism for contracting officers to structure contracts to be adaptable to contractors' innovations resulting in improvements in cost, schedule, and performance and to reward them with increased "market share" (FAR 43, 2025).

Vital to contract modification is the Government's responsibility to notify contractors of any changes to the conditions of the contract. Notification of contract changes allow contractors—when they consider that the Government has effected or may effect a change in the contract that has not been identified as such in writing and signed by the contracting officer—to notify the Government, in writing and as soon as possible, to permit evaluation of the alleged change (FAR 43.104, 2025).

Finally, contract clauses are available for use primarily in negotiated research and development or supply contracts for the acquisition of major weapon systems or principal subsystems. Further, they are used when the contracting officer anticipates that situations will arise that may result in a contractor alleging that the Government has effected changes other than those identified as such in writing and signed by the contracting officer (FAR 52.243, 2025).

In summary, the FAR currently contains the components to accomplish the aims of Concept Goal 1. The results of the research revealed that there is currently no way to engage multiple contractors on the same contract—or a portfolio contract. However, the FAR contains components which, if reconfigured, would support the implementation of a functional portfolio contract.

Separately, the research revealed that the ability to alter the terms of contracts currently exists. However, the intent of this ability is focused on being prepared to alter the contract in response to changes in supply/service demand or extraordinary contractual relief. This runs contrary to the intent of this project: designing a contracting model that can incorporate innovation <u>and</u> reward contractors that outcompete other contractors by investing in product improvement.

Requirements Documents

Manual for the Operation of the Joint Capabilities Integration and Development System

The following summarizes research into the manual for the operation of the Joint Capabilities Integration and Development System (JCIDS) and capabilities requirement document fundamentals. The focus of this research was to (1) understand the fundamentals of requirements documents and (2) determine how these documents can either enable or hinder innovation.

JCIDS operates through organizational structure and provides baseline for documentation, review, and validation of capability requirements across the Department of Defense (DoD). Validated JCIDS documents facilitate doctrine, organization, training, materiel, leadership, personnel, facility, and policy (DOTmLPF-P) changes, guide the AAF pathways, and inform planning, programming, budgeting, and execution (PPBE) processes.

Once validated, regardless of validation authority, Sponsors upload final versions of JCIDS documents and their associated memoranda into the knowledge management / decision support (KM/DS) system. This is done for archiving purposes and for visibility in the capability portfolios (Joint Chiefs of Staff [JCS], 2021).

Regarding science and technology (S&T) and innovative approaches, once proven at the appropriate technology level and S&T effort, prototype, and/or other innovative approach must align with existing capability requirements (which is the case for this research), or be supported by an analysis that makes a defendable case for a new capability.

There are two main entry points into JCIDS for S&T and innovative approaches. For evolutionary technologies that support an expeditious deployment of successful weapon system component or technology prototypes in accordance with Title 10, U.S. Code, Section 2447d, JCIDS is flexible enough to consider entry at Milestone B with a new or updated capability development document (CDD) provided there is traceability to a validated capability requirement (joint or DoD component urgent or emergent operational need, or initial capabilities document).

For disruptive, game changing technologies, such as those concepts that would be generated from the National Defense Strategy (i.e., robotics and system autonomy, miniaturization, big data, human-machine collaboration, development of new Joint Operating Concepts, etc.), there is a requirement (concept, threat informed) for the Warfighter community to determine whether it changes their CONOPS. If it does, then the appropriate entry point would be an updated capabilities based assessment (CBA) to determine what new set of missions/task/capabilities are required to fulfill a new or existing capability gap (JCS, 2021).

Consolidated Requirements Document for Search-Based Unmanned Underwater Vehicles in Support of Expeditionary Operations

Pivoting to a specific requirements document, the Consolidate Requirements Document (CRD) for Search-Based Unmanned Underwater Vehicles (UUV) in Support of Expeditionary Operations provides a comprehensive explanation of requirements for the development, production, employment, and maintenance of UUVs. A thorough review of the document revealed six key components that drive the development of capabilities and determine the degree of flexibility that contractors have, in terms of innovation.

Joint capability areas are collections of like DoD capabilities functionally grouped to support capability analysis, strategy development, investment decision making, capability portfolio management, and capabilities-based force development and operational planning. They provide a common capabilities language for use across the activities and processes of the DoD.

Tier 1	Tier 2
Joint/Maritime/Littoral Operations	Undersea Warfare
	Maritime/Littoral Expeditionary Operations
Joint Access & Access Denial Operations	Forcible Entry
	Sea Lines of Communication (SLOC) protection
	Freedom of Navigation
Stability Operations; Military Support for Stability, Security, Transition, and Reconstruction (SSTR)	Security
Joint Special Operations & Irregular Warfare	Special Reconnaissance
	Unconventional Warfare
Joint Homeland Defense	Maritime Defense
	Critical Infrastructure Protection

Figure 5. Search-Based UUV in Support of Expeditionary Operations—Joint Capabilities Areas

The family of systems (FoS) concept describes multiple system that are similar enough to be developed in support of fulfilling an operational capability gap. System(s) can be developed by a single contractor or by multiple contractors, designing to common operational requirements. This facilitates configuration control and consistency in operations and maintenance.

The threat summary describes the potential operational conditions the system can reasonably be expected to encounter, which translates into risk to mission and force. This summary drives the risk mitigations that must be considered for incorporation into system design.

The program summary describes the conceptual architecture of program management. This includes key operational system attributes and program intent for the evolution of the system, informing efforts to balance cost, schedule, and performance constraints.

Key performance parameters (KPP) establish the key aspects of performance that determine the overall operational effectiveness of the system.

6.1 Key Performance Parameters (KPPs)

The following are mandatory Key Performance Parameters (KPPs) that the UUV FoS shall meet unless otherwise noted.

Table 6-1. Key Performance Parameters

Key Performance Parameters (KPPs)	Man Portable								Lightweight								
	Threshold (T)				Objective (O)				Threshold (T)				Objective (O)				Notes
	Increment				Increment				Increment				Increment				A=Non-Complex Environment (e.g. A-1) B=Complex Environment (e.g. B-3)
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D	
P_DP_C (Bottom Targets)	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.90	0.80	0.80	0.85	0.90	0.80	0.85	0.90	0.95	All manportable increments and Lightweight Incr. A = Proud *Only*
Environment	1	1	1	2	2	2	2	2	1	2	2	2	2	2	2	2	See notes 1, 2 below
P_DP_C(Volume targets)									TBD	TBD	TBD	TBD	0.90	0.90	0.90	TBD	All manportable increments and Lightweight Incr. A = Proud *Only*
Environment									2	2	2	2	2	2	2	2	See notes 1, 2 below
P_{fa}	0.20	0.20	0.10	0.10	0.15	0.15	0.10										All manportable increments and
Environment	1	1	1	1	2	2	2										
P_{ID}	0.80	0.80	0.85	0.85	0.80	0.80	0.85										
Environment	1	1	1	2	1	2	2										
CLA drms	15				6												

(1). "Non complex Environments" environments are seabed
dimensional volume of UXO casings remains proud of the sea
rock, outcroppings or marine growth among homogenous sea
UUV operating area. In traditional MCM terminology a Type A-
(2). "Complex Environments" environments are seabed and b
units) within what would otherwise be optimal imaging rang
outcroppings, walls, or other natural obstructions, significant
percent mine case burial or concealment, and high clutter de
traditional MCM terminology, a Type C-3 or D-3 bottom would be an example of these conditions.

Tier 1 to 3 JCAs	Key Performance Parameter	JPR (check box)	Threshold	Objective
	KPP 1: Force Protection	☐	Value	Value
	KPP 2: System Survivability	☐	Value	Value
	KPP 3: Energy	☐	Value	Value
	KPP 4: Sustainment	☐	Value	Value
	KPP 4.1: Materiel Availability A_M	☐	Value	Value
	KPP 4.2: Operational Availability A_O	☐	Value	Value
	KPP 5: Joint Interoperability	☑	Value	Value
	KPP 6: Net Ready	☑	Value	Value
	KPP 7: Joint Training Interoperability	☑	Value	Value

Figure B- 8: KPP Table Format Example

Figure 6. Search-Based UUV in Support of Expeditionary Operations—KPPs

Key system attributes (KSA) establish the key measures that influence cost, schedule, and performance management for the full life of the system.

6.2 Key System Attributes

The following are mandatory KSAs for the UUV unless otherwise stated. These are summarized in Table 6-2.

Key System Attributes (KSAs)	Manportable		Lightweight		
	Threshold (T)	Objective (O)	Threshold (T)	Objective (O)	Notes
ACR (nmi²/hr) (MCM Search Mission)	0.02	0.04	0.07	0.14	$ACR_{Vehicle}$ = f(vehicle speed, sensor performance ,battery Endurance) $ACR_{Sustained}$=f(vehicle speed, sensor performance, battery endurance ,pre/post-mission analysis, turnaround time)
ACR (nmi²/hr) (IPOE Mission)			TBD	TBD	$ACR_{Vehicle}$ = f(vehicle speed, sensor performance, battery Endurance) $ACR_{Sustained}$=f(vehicle speed, sensor performance ,battery endurance, pre/post-mission analysis, turnaround time)
Contacts/Mission (ID Mission)	12	20			Contacts per Mission parameter applies only to Manportable ID mission; Assumes 8hr mission; L/W systems perform search only mission until at least Inc3.
A_O	0.80 (80% CI)	0.95 (80% CI)	0.80 (80% CI)	0.95 (80% CI)	Same assumptions for both Mods
R_o	0.80 (80% CI)	0.95 (80% CI)	0.80 (80% CI)	0.95 (80% CI)	Same assumptions for both Mods
Vehicle Weight (lbs)	160	120	1000	800	Needs to be driven by L&R system limitations (length, weight).
Payload Weight (lbs)	N/A	N/A	450	160	Needs to be driven by L&R system limitations (length, weight).
MCMTOMF (hrs)	5	2	5	2	
Threat Depth Range (ft)	10-300	10-300	10-300	10-900	Clarify environmental limitations that may affect depth capabilities.
Transit Distance to OpArea (nmi)	5	12	10	25	
Interoperability	See Notes	See Notes	See Notes	See Notes	IERs=COIN, MEDAL, EUNS - Use BULS words as a start

	Increment				Increment	Notes
	A	B	C	D	A	A-D represent capability increments
C2 Range (nmi)	0.5	1	3	3	0.5	A
Endurance (hrs) (Search Mission)	8	8	4	6	8	
P_{ID}						
Environment						
Total Ownership Cost						

Tier 1 to 3 JCAs	Key System Attribute	Threshold	Objective
	KSA 1: Reliability	Value	Value
	KSA 2: Maintainability	Value	Value
	KSA 3: O&S Costs	Value	Value
	KSA 4: Exportability	Value	Value
	KSA 5: SWaP-C		

Figure B- 9: KSA Table Format Example

Figure 7. Search-Based UUV in Support of Expeditionary Operations—KSAs

The CRD aptly summarizes its functionality by addressing the nature of the program: "The [UUV] FoS consists of small, man-portable unmanned systems for confined area operations and larger, lightweight unmanned systems for search operations in complex environments, each of which will use a common operator interface. The development of the FoS in achievable increments, or alternatively in pursuit of a next generation system, will also allow a FoS architecture to be developed while accommodating effective risk management," (Chief of Naval Operations, N957, 2012).

Thus, the JCIDS process combined with a CRD written to enable generational system evolution already exists. The responsibility to write the requirements document in a way that enables innovation integration lies with the human in the loop.

Private Industry

An article entitled *How Procurement Portfolio Management Supports the Procurement Process* discusses the concept of and benefits provided by procurement portfolio management. This term refers to the "strategic management of an organisation's procurement activities" and "promotes a holistic approach to procurement process management as it considers the organisation's overall procurement needs, goals, and strategies" (Kronos Group, 2023).

Procurement portfolio management focuses "not on individual procurement projects, but on the overall impact and value of procurement to the organization. This involves making informed decisions about resource allocation, procurement initiative prioritization, and optimizing performance and outcomes across the entire procurement profile."

Implementing portfolio management practices enable strategic alignment, risk management, and resource optimization.

"With a procurement portfolio established, organisations can match it up with business objectives and manage priorities effectively. As a procurement portfolio provides a centralized view of an organization's procurement needs and goals, aligning the procurement process with the organization's overall objectives becomes straightforward."

"With a procurement portfolio established, organizations have the potential to identify and mitigate risks and ensure supply chain resilience. A well-established procurement portfolio also provides an overview of the risks an organization could face throughout the procurement process, allowing it to formulate strategies for avoidance or mitigation."

"With a procurement portfolio established, organizations can cut costs and improve efficiency with optimal resource allocation. Since a procurement portfolio provides an extensive amount of information in a concise, condensed format, identifying opportunities for spend optimization and effective resource allocation becomes much simpler," (Kronos Group, 2023)

An article entitled "The Procurement of Strategic parts. Analysis of a Portfolio of Contracts with Suppliers Using a System Dynamics Simulation Model" investigates the employment of procurement portfolio management in the valuation of real options. Because procurement and financial managers use real options to "secure price and availability in the face of volatile world demand," portfolio valuation is "critical to option pricing models" (Marquez & Blanchar, 2004).

The above articles describe how procurement portfolio management can benefit the Government, with regard to strategic alignment, risk management (supply and price), and optimizing resource allocation. But this is only part of the solution. To shift innovation ownership to contractors, the Government must leverage the drive for competitive advantage after contract award.

The article entitled "Market Share: Understanding Competitive Advantage through Market Power" evaluates the validity of measures that relate to market share—like stability and concentration metrics—as indicators of a company's sustainable competitive advantage (Mauboussin & Callahan, 2022).

The dictionary definition of market share is "the percentage of the market for a product or service that a company supplies." (Merriam-Webster) As such, "market share is an outcome of a company’s product or service offering, distribution channels, marketing initiatives, and customer relationships” (Mauboussin & Callahan, 2022).

These are all business aspects that the contractor owns and has the power to improve. Figure 4 diagrams how companies can generate and sustain advantage.

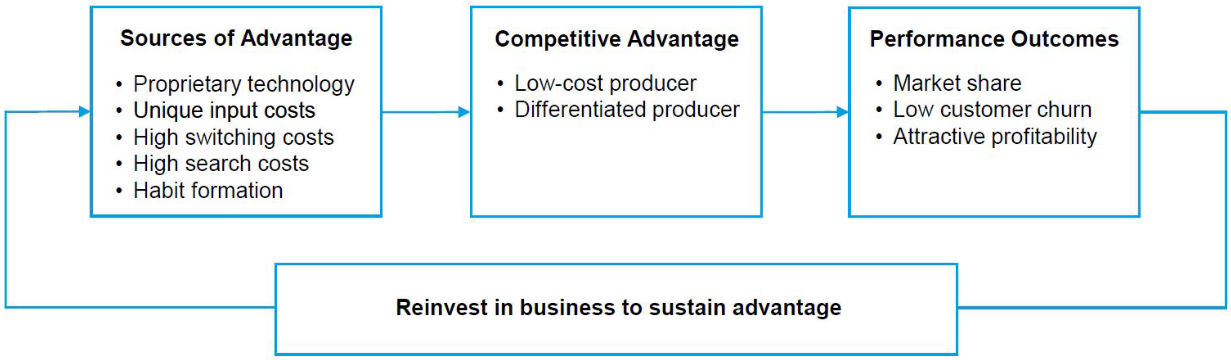


Figure 8. Traditional Competitive Strategy Analysis (Mauboussin & Callahan, 2022)

By controlling market share, the Government can establish micro-markets (portfolio contracts), which contain multiple contractors, competing to increase market share through innovation.

This takes us back to the third precept of Concept Goal 1—develop a contract able to reward contractors who innovate by increasing the quantity of supply demanded (modifying the terms of the contract)—at the expense of the competing contractors. Another way of stating this is: develop a variable, portfolio contract.

Variable, Portfolio Contract

Concept

The variable, portfolio contract (VPC) conceptually combines aspects of currently existing FAR-based contract components, to create a single contract capable of simultaneously engaging multiple contractors, in a relationship with the Government.

The VPC is fundamentally an indefinite-delivery contract, either designated as a requirements contract or indefinite-quantity contract, employing multiple award preference.

The current uniform contract format is sufficient to document contract conditions necessary for the Government to engage multiple contractors, in fulfillment of materiel capability delivery.

Bilateral (or multilateral, in the case of the VPC) contract modifications provide the mechanism for contracting officers to structure contracts to be adaptable to cost, schedule, and performance improvements/innovations by contractors and to reward them with increased market share.

The use of contract clauses provide the legal standing, dependent on collective agreement, to modify the contract when triggered by innovations.

Contracting officers must proactively communicate notifications of contract changes to ensure all parties are aware of impending changes to the contract’s conditions so that no one is surprised by market share changes.

Model

The VPC relies on the Government's ability, represented by the contracting officer, to accurately rate contractors based on their management of cost, schedule, and performance—against each other. Additionally, this rating system must be complex enough to capture the effectiveness/efficiency of cost, schedule, and performance management, but simple enough to rapidly assess these factors and update changes in comparative ratings. Further, the VPC must be able to apply a weighting system to enable decision makers to apply priorities in contractor rating. Thus, the VPC model to assess contractor rating includes the following components:

1. Rating Factor:
    a. Production Schedule (or Productivity; Schedule Factor)
    b. Production Cost (Cost Factor)
    c. Product Performance (Performance Factor)
2. Rating Factor Score: a number representing the rank of individual contractors out of the total number of contractors engaged by the Government (reverse order, i.e., worst score is "1," best score is total number of engaged contractors)
3. Rating Factor Weight: scale of 0.0–1.0; all factor weights must add up to 1.0
4. Contractor Rating: sum of individual contractor's factor score multiplied by the factor's weight (sum of all contractor ratings is 1.0)
5. Periodicity of contractor rating reevaluation (i.e., monthly, quarterly, semiannually, etc.; based on the duration of the contract)

Table 1 depicts a sample VPC Contractor Rating Calculator for a contract engaging four contractors with the Government. In this example, the Rating Factors are all weighted equally.

Table 4. VPC Contractor Rating Calculator

		Factor Weight	0.3333333	Factor Weight	0.3333333	Factor Weight	0.3333333		
Producer	Contractor Rating	Production Rate (per month)	Score (1-4)	Production Cost (per unit)	Score (1-4)	Product Performance	Score (1-4)	Producer	Market Share
A	0.23	21	2	$ 90,000.00	2	8	3	A	23%
B	0.30	30	4	$ 70,000.00	4	6	1	B	30%
C	0.20	17	1	$ 100,000.00	1	10	4	C	20%
D	0.27	25	3	$ 80,000.00	3	7	2	D	27%

Once initial contractor ratings are calculated, their decimal value is converted to a percent and these values represent the market share—or percentages of total units demanded from the individual contractors.

When individual contractors implement an innovation that improves the metrics in the cost, schedule, and performance factors (enough to alter the factor score), a contractor rating review is initiated and updates to ratings and market share are enacted followed by notifications of contract change.

Implementation Concerns

The following is a summary of discussions with a broad spectrum of Defense professionals. This includes Project Managers, Program Executive Officers, Task Group Commanders, Joint Staffers, and contracting experts. The main concerns for VPC implementation focus on mitigating risk to contractor operations, configuration management, and addressing the potential for unanticipated PPBE benefits.

Several discussions focused on the business operations of the engaged contractors. Specifically, in a system where market share is variable—i.e., the units demanded from a contractor can increase <u>and</u> decrease—how does a business engage in a way that validates its production investment?

Issues that were identified included the fact that "*industry* needs to maintain levels of production to validate infrastructure investments (people, machinery, factories, etc.)." Also, "large ramp-up/ramp-down orders are not sustainable for small businesses" (E. Hui, personal communication, March 7, 2025).

Another question that was asked focused on the potential for contractors to resist the intra-contract competition aspect of the VPC (S. Clark, personal communication, March 4, 2025). This concern appears to be rooted in the desire for businesses to maintain stability of operations and reduce uncertainty.

Another focal point centered on the concern for configuration management. Specifically, configuration management "ensures that personnel know exactly how to op test, mission plan, employ, and recover equipment" (E. Ford, personal communication, February 24, 2025).

Further, the "difficulty of maintaining a baseline (physical and logical components that make up a product) is increased with the VPC's 'micro-market'" (J. Haase, personal communication, February 25, 2025).

This is an extremely valid concern, considering that the VPC concept intentionally engages multiple contractors in the development, production, and delivery of Warfighter capability. As such, it is absolutely vital that requirements documents mandate certain common, system aspects be incorporated into the products. This ensures that no matter what contractor delivers the product—or what mixtures of product are held in inventory—the set-up, employment, operation, and maintenance are as identical as is feasible.

The last main focal point addressed the potential, positive effects of the VPC construct. The scenario posed involved the situation in which a "VPC contractor funds innovation on an existing product and that product now meets or exceeds a requirement the Government has a separate R&D contract for."

The proposed response for this scenario was that the "government should be able to reprogram the R&D funds to buy more of the improved, existing product," thus filling funding gaps in other programs (D. McDonald, personal communication, February 24, 2025).

This insightful questions addresses a key imperative of Government acquisition and procurement: How does the Government maximize positive, second-, and third-order effects through process improvement?

Methodology

This section identifies the scenario developed to test the VPC model, defines the experimental conditions, and presents a hypothesis for the VPC's performance.

Scenario

Based on the Navy's efforts to develop and employ a bi-modal—or hybrid—fleet model, the VPC will be tested in a scenario requiring the provision of unmanned systems in support of maritime domain awareness (MDA) and underwater (UW) effects. Specifically, this experiment focuses on developing a contract to provide UUVs, to scan from just beneath the surface to just above the seabed, to provide baseline operational environment awareness and change-detection for full-spectrum (from passive MDA to UW "hellscape"), underwater effects (kinetic and non-kinetic).

The UUVs provided by the VPC will be deployed in a specific geographic location (sector), for a limited period of time in support of sea denial and sea control. They are tended by unmanned surface vehicles/vessels (USV) which download sensor data and upload new tasking (providing a greater degree of autonomy to the unmanned assets of the bi-modal fleet). These USVs then transmit downloaded data to a fusion cell to feed the MDA common operational picture (COP).

Anticipating support to major combat operations (MCO), the expectation is that the UUVs will experience a high rate of attrition, due to environmental hazards and adversary actions. This necessitates the following:

1. An initial operating inventory, sufficient to cover the assigned sector
2. A reserve inventory, sufficient to reduce impacts of estimated attrition (casualty or kinetic effect-based) of operational units of action
3. Contract capacity to expand procurement of operational units of action (UoA) in the event of a greater-than-capacity (GTC) expense event (casualty or kinetic effect-based)

Capability Requirements

As addresses in the requirements document portion of the research section, producers engaged in a VPC working to (1) provide a product that meets required standards and (2) have the flexibility to invest in cost-benefit-positive innovation, requirement must be broad enough to enable unique capability solutions <u>and</u> include common design elements that facilitate uniformity of operations/maintenance for the end user.

The following requirements and common design elements seek to enable both design flexibility and uniformity of operations/maintenance:

1. Design Flexibility:
    a. Must be able to operate in the full spectrum of physical operating environmental conditions (temperature, salinity, turbidity, current, etc.)
    b. Must be able to operate from very shallow water (10 FSW) to maximum depths (as identified for MDA)
    c. Must be transportable/shippable via air, sea, rail, road safely/securely and arrive in operating condition
    d. Must be deployable based on maximum acceptable time from unpacking (from transit) to ready-for-deployment
    e. Must be deployable into the operational environment via all platforms (surface vessel, subsurface vessel, air-delivered, etc.) and man-transportable and/or lightweight
    f. Must be able to accomplish all anticipated effects-based missions:
        i. Intelligence preparation of the operational environment
        ii. Environmental (UW) change detection
        iii. Specific location/identification/mapping/targeting of critical UW infrastructure
        iv. Payload delivery of kinetic/non-kinetic effects, etc.
    g. Must be able to carry full spectrum of anticipated effects-based payloads (sensors, communications, munitions, mechanical devices, etc.)
    h. Must incorporate “scuttle” options to prevent adversary exploitation

i. Must meet minimum operational duration
j. Must meet minimum data storage capacity
k. Must be able to receive programming system upgrades
l. Must be able to receive hardware upgrades or be exchanged (swapped) at lower-than-procurement cost

2. Uniformity of Operations and Maintenance:
    a. Must be controlled on a common, user interface device
    b. Must be able to interface with an autonomous/semi-autonomous, controlling UoA (USV)
    c. Must be able to recharge via universal charger (location-agnostic: seabed-, “mothership”-based)
    d. Must be able to upload data to and interface universally with government systems [note: this potentially identifies the demand signal for a Government-procured/developed, universal data share platform]
    e. Must be serviceable by a system-agnostic field service representative (FSR), based on:
        i. Level of field maintenance capability required
        ii. Mean corrective maintenance time per operational mission
        iii. Minimum, universal repair kit available

Experiment Boundaries

UoA Quantity Requirements

1. Required operational duration (contract): 36 months (October 1, 2025–September 30, 2028)
2. 1,000 UoA operational at any given time for a 24-month period
3. Estimated attrition rate (per month): 50 UoA (5%)
4. Reserve inventory: 100 UoA
5. Total estimated quantity requirement (contract): 2,300 UoA
6. Total start-up requirement (due October 1, 2026): 1,100 UoA

Assumptions

1. Contractor production cost is equal to Government cost of procurement
2. Contractors all produce UoA that meet minimum capability requirements (product performance score of 6)
3. All contractors voluntarily adhere to requirements of VPC (including acceptance of market share changes)
4. The Government is able to engage enough contractors to meet the minimum, required production capacity of the VPC

Rules

1. Producer ratings assessed prior to contract execution and:
    a. Experiment 1: reassessed when triggered by innovation event
    b. Experiment 2: reassessed periodically (quarterly, semiannually, annually)
2. Acceptable product performance range: 6–10
3. Government cost of procurement ceiling: $110,000 per UoA (2026 dollars)

4. Inflation component applied annually (at start of fiscal year): Producer Price Index (PPI; 2.97% as of December 30, 2024)

Experiment Variables

Control Variables

1. Factor weights
2. Innovation events
3. Inflation component

Independent Variables

1. Contractor's production rate (per month)
2. Contractor's production cost
3. Contractor's product performance

Dependent Variables

1. Contractors:
    a. Factor Weight rank
    b. Contractor Rating / Market Share
    c. Initial UoA quantity demanded (out of 1,000 total)
    d. Reserve inventory UoA quantity demanded (out of 100 total)
    e. Replacement UoA quantity demanded (out of 50 monthly)
    f. Contribution to total contract procured UoA
    g. Average cost per UoA (full contract)
2. Total Contract:
    a. Total contract cost (for each scenario)
    b. Total UoA produced (for each scenario)
    c. Average cost per UoA (for each scenario)
    d. Average performance rating (for each scenario)
    e. Replacement Time (based on excess production capacity available)

Experiment Conditions

Contractor Rating Information

Table 2 identifies four contractors (Producers A, B, C, and D) engaged with the Government via a VPC, and provides the cost, schedule, and performance information used to calculate contractor rating.

Table 5. VPC Contractor Rating Information

Contractor	Production Rate (UoA, per month)	Production Cost ($, per unit)	Product Performance (1-10)
Contractor A	#	$#	#
Contractor B	#	$#	#
Contractor C	#	$#	#
Contractor D	#	$#	#

Innovation Events (Experimental Scenarios)

The following is a list of scenarios used to test the VPC:

1. Static operational conditions scenario(Control; no innovations occur during execution of the VPC)
2. Production Schedule Improvement Scenario
   a. October 1, 2026: Producer A increases productivity by 57.9%
3. Production Cost Improvement Scenario
   a. October 1, 2026: Producer C decreases production cost by 21%
4. Product Performance Improvement Scenario
   a. October 1, 2026: Producer D increases product performance by 28.6%
5. Various Factor Improvement Scenario (sequenced)
   a. January 1, 2027: Producer C increases productivity by 76.9%
   b. April 1, 2027: Producer A decreases production cost by 21.5%
   c. October 1, 2027: Producer B increases product performance by 25%
6. GTC Expense Scenario
   a. March 2028: Operational units suffer 25% casualties (250 units)

Experimental Weights

The following is a list of weights applied to each of the scenarios identified above:

1. Even weight
2. Productivity-weighted
3. Cost-weighted
4. Performance-weighted

A foundational component of this research centers on reducing supply risk through the employment of a portfolio contract. As such, preference in weighting is given to the productivity (schedule) factor. For this reason, productivity is not given the lowest weight for any of the scenarios, as depicted in Table 3.

Table 6. VPC Experiment Weighting System

Contractor Rating Factor Weights	Production Rate (per month)	Production Cost (per unit)	Product Performance
Even	0.333	0.333	0.333
Schedule	0.5	0.3	0.2
Cost	0.3	0.5	0.2
Performance	0.3	0.2	0.5

VPC Model Evaluation Hypotheses

1. Implementation of a VPC will create an environment in which innovating contractors are rewarded with increased market share
2. Innovations will accomplish the following:
   a. Reduce total contract cost
   b. Improve average UoA performance
3. Productivity-weighted VPC will yield greatest reduction to supply risk

VPC Model Evaluation and Results

Model Evaluation Tool

The evaluation tool was built on the Microsoft Office Excel application. The tool consisted of three separate sheets within a single workbook.

The first sheet, entitled Data Input, provided the following functions (Tables 2 and 3 data inputted for each experiment):

1. Contractor factor input table, including:
    a. Contractor Production Rate
    b. Contractor Production Cost
    c. Contractor Product Performance
    d. Contractor Factor Score
    e. Factor Weight
2. Calculated the contractor rating and market share
3. Calculated and depicted market share of total UoA demanded from each contractor, broken down into the following categories:
    a. Operational UoA: 1,000 units
    b. Initial Reserve UoA: 100 units
    c. Estimated Attrition Replacement Rate (EARR): 50 units per month
4. Calculated maximum VPC production capacity
5. Calculated VPC EARR surplus/deficit

The second sheet, entitled Schedule-Cost, provided the following functions:

1. Calculated and displayed the VPC costs, broken down by:
    a. Month
    b. Year
    c. Total Contract Cost
    d. Total contract average cost per UoA
    e. Contract Cost per Contractor
2. Calculated and displayed UoA procured by the VPC, broken down by:
    a. Contractor per month
    b. Contractor per year
    c. UoA producer per contractor
    d. Total produced by contract
3. Calculated and displayed Contractor Rating, broken down by:
    a. Final contractor rating (at contract termination, or post-Contractor Rating reevaluation)
    b. Average contractor rating of the contract
4. Calculated and displayed Product Performance, broken down by:
    a. Final contractor Product Performance (at contract termination, or post-Contractor Rating reevaluation)
    b. Average product performance of the contract

The third sheet, entitled Data Analysis, depicted consolidated results for each scenario as well as comparisons of data to the Control. The result categories depicted included:

1. Contractor Rating
2. Total Cost of Contract
3. Total UoA procured
4. Average cost per UoA for total contract
5. Average Product Performance for total contract

The comparison between the Control and the individual innovation scenarios, included:

1. Change in average cost per UoA for total contract
2. Change in total cost of contract
3. Change in average product performance for total contract

For access to the model evaluation tool and/or raw data, contact the author.

Model Evaluation Results

The model evaluation consisted of 120 experiments run, broken into four, 30-scenario batches. These batches each employed one of the four Contractor Rating Factor Weights.

Data Analysis

Tables 4, 5, 6, and 7 display the analytical results of the four batches of experiments (contact the author for raw data).

Tables 4, 5, 6, and 7. VPC Experiment Summaries (Weights: Even, Schedule, Cost, and Performance; respectively)

Even-Weighted	Change in Average Contract Performance (from Control)									Market Share Analysis	
	Total Cost	% Change	Std Dev	Contract Capacity (UoA per month)	% Change	Std Dev	Performance Rating	% Change	Std Dev	Average Contractor Minimum (%)	Average Contractor Maximum (%)
Control Scenario	$ 177,692,395.52			99.8367			7.8380			18.0783%	33.3148%
Schedule Improvement Scenario	$ 177,828,354.84	0.0951%	0.0108	113.8311	14.0545%	0.0194	7.8640	0.3605%	0.0147	Minimum UoA Demanded (individual Contractor)	Maximum UoA Demanded (individual Contractor)
Cost Improvement Scenario	$ 171,427,410.33	-3.5266%	0.0069	99.8367			7.8830	0.5918%	0.0079	423.3333	780.1222
Performance Improvement Scenario	$ 177,187,468.71	-0.3008%	0.0051	99.8367			8.4403	7.6892%	0.0204	Average Total Contract Cost Percent Change	$ 180,748,442.03 1.7199%
Various Improvement Scenario	$ 183,385,819.65	4.1534%	0.0630	117.9338	18.1613%	0.0375	8.4583	7.8660%	0.0207	Standard Deviation Avg Contract Perfromance Rating	0.0108 8.0536
G-T-C Scenario	$ 196,969,203.12	10.8473%	0.0022	99.8367			7.8380	0.0000%	0.0000	Percent Change Standard Deviation	2.7508% 0.0059

Schedule-Weighted	Change in Average Contract Performance (from Control)									Market Share Analysis	
	Total Cost	% Change	Std Dev	Contract Capacity (UoA per month)	% Change	Std Dev	Performance Rating	% Change	Std Dev	Average Contractor Minimum (%)	Average Contractor Maximum (%)
Control Scenario	$ 176,800,603.90			99.8367			7.7450			15.8598%	36.0685%
Schedule Improvement Scenario	$ 177,004,542.72	0.1662%	0.0163	113.9662	14.1896%	0.0198	7.7847	0.5740%	0.0225	Minimum UoA Demanded (individual Contractor)	Maximum UoA Demanded (individual Contractor)
Cost Improvement Scenario	$ 171,049,281.79	-3.2468%	0.0086	99.8367			7.7867	0.5570%	0.0072	371.3833	844.6040
Performance Improvement Scenario	$ 176,490,158.02	-0.1884%	0.0031	99.8367			8.3503	7.8135%	0.0119	Average Total Contract Cost Percent Change	$ 180,068,374.38 1.8483%
Various Improvement Scenario	$ 183,050,204.91	4.4981%	0.0641	118.1132	18.3407%	0.0378	8.3950	8.3669%	0.0256	Standard Deviation Avg Contract Perfromance Rating	0.0116 7.9678
G-T-C Scenario	$ 196,015,454.96	10.8655%	0.0013	99.8367			7.7450	0.0000%	0.0000	Percent Change Standard Deviation	2.8764% 0.0079

Cost-Weighted	Change in Average Contract Performance (from Control)									Market Share Analysis	
	Total Cost	% Change	Std Dev	Contract Capacity (UoA per month)	% Change	Std Dev	Performance Rating	% Change	Std Dev	Average Contractor Minimum (%)	Average Contractor Maximum (%)
Control Scenario	$ 174,671,674.91			99.8367			7.7297			15.6114%	35.8903%
Schedule Improvement Scenario	$ 174,794,038.20	0.0896%	0.0100	113.9662	14.1896%	0.0198	7.7553	0.3575%	0.0135	Minimum UoA Demanded (individual Contractor)	Maximum UoA Demanded (individual Contractor)
Cost Improvement Scenario	$ 168,876,628.03	-3.3030%	0.0064	99.8367			7.7993	0.9292%	0.0119	365.5667	840.4307
Performance Improvement Scenario	$ 174,361,229.02	-0.1882%	0.0031	99.8367			8.3397	7.8942%	0.0121	Average Total Contract Cost Percent Change	$ 177,918,198.72 1.8586%
Various Improvement Scenario	$ 181,151,016.49	4.6471%	0.0622	118.1132	18.3407%	0.0378	8.3533	8.0335%	0.0222	Standard Deviation Avg Contract Perfromance Rating	0.0108 7.9512
G-T-C Scenario	$ 193,654,605.65	10.8655%	0.0013	99.8367			7.7297	0.0000%	0.0000	Percent Change Standard Deviation	2.8656% 0.0067

Performance-Weighted	Change in Average Contract Performance (from Control)									Market Share Analysis	
	Total Cost	% Change	Std Dev	Contract Capacity (UoA per month)	% Change	Std Dev	Performance Rating	% Change	Std Dev	Average Contractor Minimum (%)	Average Contractor Maximum (%)
Control Scenario	$ 181,202,970.06			99.8367			7.9670			19.6078%	31.6780%
Schedule Improvement Scenario	$ 181,325,333.34	0.0774%	0.0095	113.9662	14.1896%	0.0198	7.9917	0.3319%	0.0131	Minimum UoA Demanded (individual Contractor)	Maximum UoA Demanded (individual Contractor)
Cost Improvement Scenario	$ 174,333,164.95	-3.8060%	0.0089	99.8367			7.9937	0.3423%	0.0046	459.1500	741.7933
Performance Improvement Scenario	$ 180,408,592.13	-0.4610%	0.0076	99.8367			8.6193	8.1629%	0.0153	Average Total Contract Cost Percent Change	$ 183,984,233.70 1.5349%
Various Improvement Scenario	$ 185,817,251.18	3.4968%	0.0635	118.1132	18.3407%	0.0378	8.5770	7.5908%	0.0238	Standard Deviation Avg Contract Perfromance Rating	0.0106 8.1859
G-T-C Scenario	$ 200,818,090.53	10.8260%	0.0032	99.8367			7.9670	0.0000%	0.0000	Percent Change Standard Deviation	2.7481% 0.0045

The following is a summary of observations from the model evaluation.

VPC Model Averages

1. Average, total contract cost increase: 1.7404%
2. Average, total UoA performance increase: 2.8102%
3. Average production capacity increase (translates to reduced supply risk):
    a. Schedule Improvement Scenario: 14.1188%
    b. Various Improvement Scenario: 18.2615%
4. Average Market Share Increase/Decrease per innovation:
    a. Average Market Share Increase: 22.986%
    b. Average Market Share Decrease: −13.682%
5. Theoretical, Minimum Contract Value (worst-case scenario):
    a. $23,030,122.16 (13.4348% market share) [Based on the following: Minimum total contract cost observed ($171,421,621.27), minimum average cost per UoA @ 2,300 procured UoA ($74,531.14), and minimum observed UoA procured by individual contractor (309)]

Individual Scenario Observations

1. Performance Scenarios—independent of weighting—result in the most reliable decrease in average, total contract cost
2. Cost Scenarios—independent of weighting—result in the most reliable increase in average, performance rating
3. Various Scenarios, followed closely by Performance Scenarios—independent of weighting—result in the highest increase in average performance rating
4. Cost Scenarios—independent of weighting—result in the highest decrease in average, total contract cost
5. Performance-Weighted Scenarios resulted in the highest minimum Contractor Market Share (19.6078%)
6. Schedule-Weighted Scenarios resulted in the highest maximum Contractor Market Share (36.0685%)
7. Cost-Weighted Scenarios resulted in the lowest minimum Contractor Market Share (15.6114%)
8. Performance-Weighted Scenarios resulted in the lowest maximum Contractor Market Share (31.6780%)
9. Minimum Contracted UoA
    a. Even-Weighted: 343.3330
    b. Schedule-Weighted: 331.0000
    c. Cost-Weighted: 309.000
    d. Performance-Weighted: 368.0000

Overall Cost, Performance, and Quantity-Demanded Observation

1. Performance-Weighted VPC demonstrated smallest average, total contract cost increase (+1.5349%; std dev: 0.016)
2. Cost-Weighted VPC demonstrated largest increase in average UoA performance rating (+2.8656%: std dev: 0.0067)

3. Minimum procured UoA (single contractor) was 309.0000, across all 120 experiments run

Conclusion and Recommendations for Further Research

Conclusion

Analysis of VPC Model Evaluation Hypotheses

1. VPC model implementation demonstrated consistent increased market share for innovators
2. VPC model's incorporation of innovations failed to demonstrate reduced, average, total contract costs
3. VPC model implementation demonstrated improved average UoA performance
4. Productivity-weighted VPC model yielded the greatest, individual reduction to supply risk—via 14.1188% increase in contract, production capacity

The model evaluation demonstrated the VPC's ability to accomplish the following:

1. Engage enough contractors to meet minimum production capacity requirements **and** provide excess capacity to respond to unanticipated spikes in procurement demand (based on GTC scenarios) thereby reducing supply risk
2. Create a competitive environment which encourages R&D investment, solely borne by the contractor
3. Effectively restructures contractor market share in response to cost, schedule, and/or performance improvements (based on reevaluated contractor ratings) thereby incentivizing innovation
4. Provides substantial incentive for all contractors to participate in the VPC, based on the theoretical, minimum contract value
5. Minimize increases in—or, in some cases, decrease—average total contract cost, while increasing average UoA product performance

Recommendations for Further Action and Research

1. Develop a uniform contract format that is structured to simultaneously engage multiple contractors.
2. Adapt the recommended requirements documents to facilitate design flexibility and uniformity in operations and maintenance (Methodology: Capability Requirements).
3. Determine the feasibility of and process for reprogramming funding, in accordance with PPBE reform guidance, in the event that VPC-based innovations result in satisfaction of R&D objectives and efforts.
4. Implement the VPC as soon as practical to ensure timely, risk-mitigated delivery of high-volume, high attrition capability for the future, maritime fight.

References

Chief of Naval Operations, Expeditionary Combat Branch (N957). (2012). *Consolidated requirements document for search-based unmanned underwater vehicles in support of expeditionary operations*. U.S. Navy.

Commission on Planning, Programming, Budgeting, and Execution Reform. (2024, March 6). *Defense resourcing for the future: Final report*.

FAR 16.5, Infinite-Delivery Contracts (2025).

FAR 16.503, Requirements Contracts (2025).

FAR 16.504, Indefinite-Quantity Contracts (2025).

FAR 15.204, Uniform Contract Format (2025).

FAR 43, Contract Modifications (2025).

FAR 43.104, Notification of Contract Changes (2025).

FAR 52.243, Contract Clauses (2025).

Joint Chiefs of Staff. (2021, October 30). *Charter of the Joint Requirements Oversight Council and implementation of the joint capabilities integration and development system* (CJCSI 5123.01I).

Kronos Group. (2023, July 2). *How procurement portfolio management supports the procurement process*. https://kronosgroup.eu/2023/07/21/why-use-procurement-portfolio-management/

Marquez, A. C., & Blanchar, C. (2004). The procurement of strategic parts. Analysis of a portfolio of contracts with suppliers using a system dynamics simulation model. *International Journal of Production Economics*, *88*(1), 29–49.

Mauboussin, M. J., & Callahan, D. (2022). Market share: Understanding competitive advantage through market power. *Morgan Stanley: Counterpoint Global Insights*.

Zsidisin, G. A. (2003). A grounded definition of supply risk. *Journal of Purchasing and Supply Management, 9*(5–6), 217–224.

THIS PAGE INTENTIONALLY LEFT BLANK

Computer Simulation in the Acquisition Life Cycle using the WRENCH Simulation Model

Dr. Susan K. Aros—is a Research Assistant Professor in the Naval Postgraduate School's Department of Defense Management and Director of NPS's Center for Modeling Human Behavior. She holds a PhD in Information, Risk, and Operations Management; master's degrees in Operations Research and Industrial Engineering, and in Spiritual Formation and Soul Care; and a bachelor's degree in psychology. She has published research in the areas of civil security operations; use of intermediate force capabilities; disaster response operations; inter-organizational communication; production planning and scheduling; and RFID use in remanufacturing. [skaros@nps.edu]

Dr. Rene G. Rendon—is an associate professor in the Department of Defense Management at NPS. A retired Air Force contracting officer, he served as a warranted contracting officer for major weapons system programs such as the Peacekeeper ICBM, F-22 Advanced Tactical Fighter, Space-Based Infrared Satellite program, and the Evolved Expendable Launch Vehicle program. He was presented with the Air Force Outstanding Officer in Contracting Award, and he has received the NPS Hamming Teaching Excellence Award and the Graduate School of Business and Public Policy Research Excellence Award. [rgrendon@nps.edu]

CPT Dajah E. Treece—is a U.S. Army Captain and native of Richmond Hill, GA. She graduated and commissioned from the United States Military Academy at West Point in 2015 with a Bachelor of Science in Mathematical Studies. CPT Treece obtained her graduate degree in Applied Mathematics from NPS. Upon receiving her graduate degree, she returned to West Point where she currently serves as a math instructor. [dajah.treece@westpoint.edu]

Abstract

Computer simulation can be used throughout the defense acquisition life cycle in order to conduct analysis of alternatives and evaluate materiel solutions; assess risk reduction efforts; and aid in development test and evaluation activities. This use of simulation can be an innovative way for transitioning emerging technologies from research and development to defense acquisition programs of record, thereby helping the program successfully cross the valley of death. In this paper, we discuss how WRENCH, a computer simulation software developed at the Naval Postgraduate School's Center for Modeling Human Behavior, could be used in the acquisition life cycle to test the effectiveness of different proposed or prototype Intermediate Force Capability (IFC) weapons configurations. WRENCH simulates a security force (SF) managing a potentially hostile crowd, enabling the exploration of potential outcomes when the SF has IFC weapons of different configurations available for use. In WRENCH, different scenarios can be tested, providing outputs to inform a variety of metrics of interest. Here we demonstrate using WRENCH to explore the effectiveness of different Active Denial Device configurations on the achievement of mission objectives in a compound defense civil security scenario, discussing an experiment, and analyzing results. We then discuss implications for acquisition and future work.

Keywords: defense acquisition, acquisition life cycle, computer simulation modeling, intermediate force capabilities, non-lethal weapons

Introduction

Assessing performance characteristics and the potential effectiveness of emerging technologies during the acquisition life cycle is challenging, particularly when physical testing of the technology is dangerous or cost-prohibitive. Computer simulation provides a method of testing that can reduce testing costs and risk. In addition, it enables testing of products at any stage in the development process from initial concept to final product. Use of simulation modeling can also expand what is possible in testing, whether the restriction on what's possible is due to high testing costs or due to the high risks of testing, such as the risk of harm to humans. In this paper, we address the use of computer simulation models tools to assess, test, and evaluate something that is being acquired. Alternate uses of M&S in acquisition may include the simulation of the acquisition process itself, as addressed in Wirthlin et al. (2011), and simulations or simulation models that are part of what is being acquired such as training simulations, as addressed in Vila (2010).

In order to provide a context for our discussion of how simulation capabilities can be useful in the acquisition life cycle, we focus on the acquisition of Intermediate Force Capability (IFC) non-lethal weapons (NLWs). Due to the fact that "NLWs and simulants carry a very real risk of permanent damage to [human] subjects" (Mezzacappa, 2014), this is a prime context for the use of simulation modeling. In their article on the importance of characterizing the human effects in NLW acquisition, Burgei et al. (2015) state that characterization of these effects is critical, as warfighters face complex engagement scenarios. The authors state "the warfighters must have confidence in the effectiveness of a NLW and understand the risk of adverse effects." The characterization of human effects in NLWs is guiding the development of these weapons in the earliest stages of the acquisition process, focusing first on warfighter needs as expressed by combat developers. The authors also argue that continually improving the human effects characterization process is key to improving NLWs (2015).

Even with the use of computer simulation, it is particularly difficult to assess the potential effectiveness of weapon systems for which non-kinetic effects can be as important as kinetic effects, and effectiveness measures pertain to human behavioral responses. In this research we explore the use of recent computer simulation modeling advances in human behavior modeling, as evidenced in the WRENCH simulation model developed at the Center for Modeling Human Behavior in the Naval Postgraduate School. We present an experiment conducted using WRENCH to demonstrate how simulation can provide insights into the relative effectiveness of different NLW configurations on the management of a dynamic, potentially-hostile crowd. Experimental results discussed here give a brief glimpse into the possibilities of using a simulation model such as WRENCH in the acquisition process.

The paper proceeds as follows. We first identify key phases in the acquisition life cycle, and then provide a brief literature review. Following that, we provide an overview of the WRENCH simulation model and describe an experiment and results analysis that demonstrates some of the capabilities of WRENCH for assessing the effectiveness of different IFC weapons. We then discuss how WRENCH could be used during different phases of the acquisition life cycle, and conclude with recommendations for using WRENCH as an innovative tool in transitioning emerging technologies from R&D to a DoD program of record and thus successfully crossing the "valley of death."

Background: The Acquisition Life Cycle

Here we provide relevant highlights of the acquisition life cycle, establishing the framework for our discussion on how modeling and simulation (M&S) can be used in the acquisition life cycle. We focus specifically on the Major Capability Acquisition (MCA) pathway, which is used to acquire and modernize military unique programs that provide enduring capability where there is a need to use a structured acquisition life cycle approach for analyzing, designing, developing, integrating, testing, evaluating, producing, and supporting the weapon system or complex capability (DoD, 2022).

The first life-cycle phase in the MCA pathway is the *Material Solutions Analysis (MSA)* phase. In this phase, activities to choose the product to be acquired (the material solution) are conducted. These activities include an analysis of alternatives (AoA). It is in this phase where validated capability gaps are translated into system-specific requirements, and planning is conducted to support an acquisition strategy for the product.

The next life cycle phase is the *Technology Maturation and Risk Reduction (TMRR)* phase. The purpose of the TMRR phase is to sufficiently reduce technology, engineering, integration and life-cycle cost risk so that the program can advance to the next phase in the life cycle.

The *Engineering and Manufacturing Development (EMD)* life cycle phase includes the development, building, testing, and evaluating of the materiel solution, to verify that all operational and implied requirements have been met, and to support production, deployment and sustainment decisions. It is during EMD that developmental testing and evaluation (DT&E) activities are conducted to provide hardware and software feedback to the program manager on the progress of the design process and on the product's compliance with contractual requirements, effective combat capability, and the ability to achieve key performance parameters (KPPs) and key system attributes (KSAs). It is also during the EMD phase that operational test and evaluation (OT&E) will be conducted to provide initial assessments of operational effectiveness, suitability, survivability, and the ability to satisfy KPPs and KSAs. The successful completion of EMD life-cycle activities supports the decision to transition to the production and deployment phase.

The *Production and Deployment (P&D)* phase includes the activities needed to produce the product (e.g., weapon system) and deploy it to operational units. These activities include completing DT&E and initial OT&E. The acquisition may also include the production of low-rate initial production units to be used in initial OT&E activities.

The *Operation and Support (O&S)* phase is the final phase of the acquisition life cycle for the MCA acquisition pathway. During this phase, activities related to operating and supporting the newly acquired weapon system are performed. These activities are in support of sustainment of the weapon system and disposal of the system after it is removed from inventory.

Literature Review

In this section, we provide a brief look at examples from the literature pertaining to the use of simulation modeling in the assessment, testing, and evaluating activities within the acquisition life cycle, drawing from real-world defense acquisition programs, followed by a discussion of the use of M&S in product development iterative cycle activities. We also discuss literature pertaining to the simulation of IFC weapons.

M&S Use in the Acquisition Life Cycle

An example of using modeling and simulation in the requirements determination process (e.g., MSA phase) can be seen in the Army's Mechanized Infantry Combat Vehicle (XM30) program. The XM30 is the Army's planned solution to maneuver warfighters on the battlefield to advantageous positions for close combat. This vehicle is expected to allow for crewed or remote operation. The Army developed the vehicle's requirements using modeling and simulation and was informed by digital concepts from different contractors during the product development phase (GAO, 2024).

An example of using M&S activities in system design (e.g., EMD) can be seen in the Army's Extended Range Cannon Artillery (ERCA) program. ERCA is part of the Army's long-range precision fires portfolio of programs. The acquisition program includes an upgrade to the M109 self-propelled howitzer that will improve lethality, range, and reliability. It will also add armament, electrical systems, and other upgrades to the existing vehicle. The Army used M&S in its iterative product development approach (GAO, 2024).

An example of M&S in reducing risk in manufacturing and testing (e.g., EMD) can be seen in the Navy's Hypersonic Air-Launched Offensive Anti-Surface Warfare Weapon System (HALO) program. The Navy's HALO acquisition program focuses on developing an anti-ship missile. HALO will address long-term capability needs for longer-range missiles with increased survivability to target heavily defended ships from near-peer competitors. The HALO program plans to leverage M&S to help address the challenge of the limited manufacturing industrial capacity to serve multiple hypersonic programs. M&S will be used by HALO contractors and their subcontractors to identify potential choke points in the manufacturing process. The program plans on using M&S in ground and flight testing in other related hypersonic programs (GAO, 2024).

Simulating Intermediate Force Capabilities

The rapid evolution of modern conflicts has highlighted the critical need for Intermediate Force Capabilities (IFCs) as a necessary element of modern military strategy. NLWs, in particular, provide options in force escalation and enable military units to disperse crowds, disable threats, and enhance force protection with reduced collateral damage (Grocholski et al., 2022). These capabilities are especially vital in addressing hybrid threats, gray zone conflicts, and unconventional warfare. In recognizing these capabilities, NATO's Military Committee has actively sought to refine IFC applications for mobility and counter mobility threats, especially in population dense environments (Afara et al., 2024).

To assess and analyze the tactical and strategic effects of IFCs, both NATO and the DoD have increasingly turned to agent-based simulation modeling. Early examples are the incorporation of unspecified NLW systems in the ModSAF simulation environment (Peters et al.,1998) and the modeling of a Long Range Acoustic Device (LRAD) in the COMBATXXI combat simulation (Grimes, 2005).

A more recent example is Gray (2017), who used an agent-based simulation model coded in Pythagoras to study the effects of a U.S. Marine patrol using a marking, blunt trauma NLW when moving through a civilian area. And just recently, Afara et al. (2024) used the Map-Aware Non-Uniform Automata (MANA), a New Zealand developed simulation tool, to study IFCs in urban mobility and counter mobility scenarios, modeling the use of "directed energy (DE) weapons such as acoustic hailers, laser warning devices, as well as microwave, millimeter wave or radio-frequencies devices" to address people blocking the advancement of a military vehicle or convoy.

Using WRENCH to Simulate ADT Effectiveness of Active Denial Technology

In this section we provide a brief overview of the WRENCH simulation model. We then discuss active denial technology (ADT), how an ADT device is modeled in WRENCH, and describe the design of an experiment of ADT effectiveness using WRENCH.

Brief Overview of WRENCH

WRENCH is an agent-based simulation model, coded in NetLogo that simulates a security force (SF) engaged in civil security operations, addressing potential threats through the use of non-lethal and lethal weapons. The mission scenario currently modeled in WRENCH is compound defense, where the SF is comprised of stationary gate guards and mobile patrol squads that can come to the aid of the gate guards during active defense of the compound. Figure 1 provides a snapshot of the compound area as depicted in WRENCH, showing the compound in the center with three designated entry points, roadways, other buildings, and also people, guards, and patrol vehicles magnified in the view for better visibility.

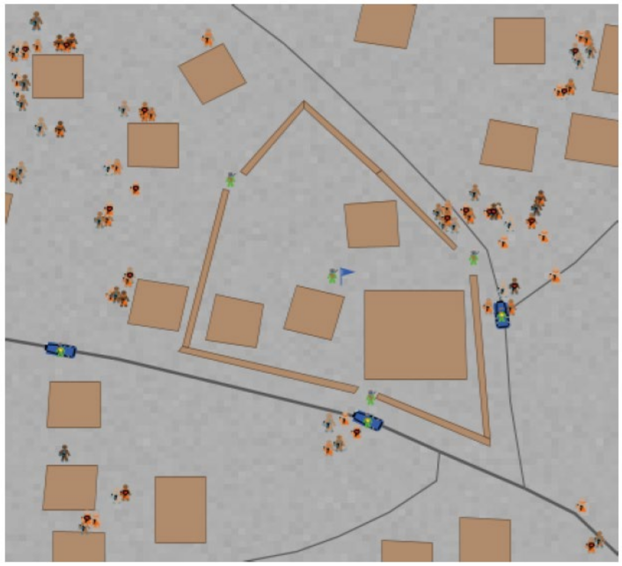


Figure 1. Snapshot of Compound and Other Simulated Elements in WRENCH (Aros & McDonald, 2023a)

Each person in the population, SF member, and patrol vehicle is modeled as an individual agent, such that each can interact with the environment and other agents autonomously. Individual agents can also be in groups (SF agents grouped within the command structure, and people agents in family or social groups). WRENCH models dynamic details for each agent such as its emotions, beliefs, needs, objective, physical state, and group influences, that will affect its cognitive decision-making and behavior.

Within WRENCH, the simulation user can specify a variety of characteristics about the SF and how they operate, the most relevant for this paper being the IFC(s) available for use, the tactical rules of engagement, and the SF's inherent stance toward the population, all of which will be explained in more detail in the experimental design section below. The user can also specify quite a few different aspects of the population, which will also be explained further within the experimental design section.

WRENCH runs with a 1-second simulated time-step, and is typically run for minutes to a few hours of simulated time in order to capture in detail the rapid changes that can occur with a potentially hostile crowd. WRENCH can be run using an interactive mode, where the emergent changes can be observed over time, or in a "headless" mode that enables large-scale experimentation of a wide variety of settings, producing a wealth of data for analysis. More details about WRENCH can be found in Aros et al. (2021) and Aros and McDonald (2023b).

Overview of Active Denial Technology

An active denial technology system is a directed energy weapon "that uses non-ionizing millimeter wave radiation to heat moisture just below the skin's surface, creating a sensation of heat" (Buch & Mitchell, 2013). Because this directed energy penetrates only a few millimeters or less into human tissue, its primary effect is limited to surface heating (Wang et al., 2020). Two active denial system (ADS) were produced under the DoD under the Advanced Concept Technology Demonstration Program. System 1 is "mounted on a modified High Mobility Multi-Purpose Wheeled Vehicle (HMMWV); and System 2 is a self-contained, box shaped model that is transportable via tactical vehicles larger than a HMMWV" (Buch & Mitchell, 2013). An ADS can project the millimeter wave beam over long distances. Safety concerns regarding overexposure are addressed by automatic shut-off mechanisms that deactivate the beam as soon as the trigger is released or when a pre-set time expires. Additionally, a laser rangefinder adjusts the output power according to the distance of the target, ensuring that safety thresholds are not exceeded (LeVine, 2009).

The ADS garnered significant media attention during and after its initial deployment, with reports discussing both its benefits and drawbacks. Some positive reported aspects of the ADS system includes the potential to minimize civilian casualties, effectively disperse large crowds, and limit collateral damage. However, negative reports on the ADS focused on the system's capacity to cause pain from far distances and numerous unforeseen and untested health risks (Buch & Mitchell, 2013).

Simulating Active Denial Technology in WRENCH

The WRENCH simulation software includes a weapons database that provides detailed specifications of various non-lethal and lethal weapons such as each weapon's range, the size and shape of the impact zone, whether it is designed for use against people or equipment, and, if it is designed for people, whether it is designed to use on a single person or multiple people in a single firing. Impacts of weapons on humans are categorized into one of seven severity levels: Psych-impact levels 1 through 3 (representing mild, moderate, or high psychological effects), Pain-impact levels 1 through 3 (representing mild/transient pain/injury, significant injury, or severe injury), or death. Any significant physical injuries or effects that alter their movement capabilities are also explicitly modeled, allowing for different patterns of impact to be modeled depending on the type of weapon.

In WRENCH, active denial devices (ADDs) are modeled as having a broader cone-shaped impact zone (possibly hitting multiple people) or a very narrow impact cone (hitting one person), with a long possible range of use, and with possible resulting impact levels ranging from Psych-2 through Pain-1. For this experiment, different ADD configurations were tested that combined differences in breadth of the impact cone, the ability to use the ADD at different power/impact levels. All but one of the tested ADD configurations were assumed to be mounted at the compound gates, one per gate, while one configuration was assumed to be hand-held; the hand-held option had a lower max range, a lower max severity level, and could be carried by a patrol squad member and a gate guard. The no-ADD case is termed "voice only" because the SF members can, in all cases, use their voice in a limited range to address hostilities. And, although the user can specify in WRENCH that the SF can have multiple different types of weapons available, this experiment limited their weapons to only the specified ADD and their voice.

A summary of the four different ADD configurations is given in Table 1. The rows specify the specific characteristics of each different numbered ADD configuration, and the columns distinguish differences when the given ADD is used at different severity levels. We emphasize that these are hypothetical weapons configurations designed to demonstrate the use of WRENCH for comparing the relative effectiveness of different weapons configurations.

Table 1. ADD Configurations Tested (Treece, 2024)

	ADD-1 (Mounted)	ADD-1 (Mounted)	ADD-1 (Mounted)
Range	1000 meters	1000 meters	300 meters
Cone Degree	30	20	10
Impact Level	Psych-2	Psych-3	Pain-1
Target Type	Multiple People	Multiple People	Multiple People
	ADD-2 (Mounted)	ADD-2 (Mounted)	ADD-2 (Mounted)
Range	1000 meters	1000 meters	300 meters
Cone Degree	15	10	5
Impact Level	Psych-2	Psych-3	Pain-1
Target Type	Multiple People	Multiple People	Multiple People
	ADD-3 (Mounted)	ADD-3 (Mounted)	ADD-3 (Mounted)
Range	1000 meters	1000 meters	300 meters
Cone Degree	0.1	0.1	0.1
Impact Level	Psych-2	Psych-3	Pain-1
Target Type	Single Person	Single Person	Single Person
	ADD-4 (Handheld)	ADD-4 (Handheld)	
Range	300 meters	300 meters	
Cone Degree	0.1	0.1	
Impact Level	Psych-2	Psych-3	
Target Type	Single Person	Single Person	

Experiment Design

To assess the relative effects of different active denial technology, a simulation experiment was conducted within WRENCH. The experiment focused on evaluating the ADD configurations just described. We also varied several additional parameters (i.e., factors) in the experiment in order to explore whether the relative effectiveness of the weapons configurations could differ when characteristics of the population and the SF were different. A summary of the experimental design parameters and levels is provided in Table 2, with discussion of each provided below.

Table 2. Experimental Parameters and Levels (Treece, 2024)

Parameter (WRENCH)	Levels (options within parameters)
Rules of Engagement (ROE) ruleset	{IH, LH, LD}
Intermediate Force Capabilities (IFC)	{ADD1, ADD2, ADD3, ADD4, Voice-Only}
Security Force Stance	{Nurturing, Cautious, Repressive}
Population Scenario	{Protest_SIG, Protest_indiv, Market}

Within WRENCH, the tactical rules of engagement are highly customizable, detailed sets of rules. The ROEs specify the basis of threat assessments (individual hostility (IH), locational hostility of clusters of people in a small area (LH), or the density of people in a small area (LD)) and the prioritization of areas and hostility levels to address first, among other details. For this experiment, the ROEs were designed to be identical, except for differing in the basis of the threat assessment. As for the available IFCs, each different ADD configuration was tested, as was a “no ADD case” where the SF were only able to use their voices. (Note that the SF members were able to use their voice in addition to the specified ADD under each ADD option as well.) The SF stance parameter provides a way to specify how the SF will interpret the observed behaviors of the people.

In WRENCH, the SF members observe the behaviors of the people and deduce a range of likely hostility levels for each person based on these observations. The Stance then specifies what level of hostility, within the range deduced from observed behavior of each person, the SF will respond to. Under the Nurturing Stance, the SF “assumes the best,” or lowest hostility level in the range; under the Repressive Stance, the SF "assumes the worst," or highest, hostility in the range; and under the Cautious Stance, they assume the mid-range hostility level.

The configurable population characteristics within WRENCH include, but are not limited to, the population demographics, distribution of initial objectives across the people, initial hostility levels, how many people "arrive" in family groups, of what size ranges, and how many people "arrive" in social groups, of what size ranges, and what basis of social group identification. In order to reduce the number of parameters and levels required to test the ADD effects on different populations, we adopted the three sample population designs first introduced in Aros & McDonald (2023b):

- Market – with a higher percentage of children and families, and fewer adults beginning with an objective to protest, and lower average initial hostility levels
- Protest – with a lower percentage of children and families, and most adults beginning with an objective to protest, with a small number of adults beginning with the objective to Attack (invade), and higher average initial hostility levels, with
- two Protest sub-types: "individual" where all people show up as individuals (except children are with mothers), and "SIG" where many people show up in social groups

The experiment conducted was a full factorial experiment across the parameters and levels summarized in Table 2, testing every possible combination of levels across parameters. This experiment approach, while inefficient, has the advantage of allowing the separation of data into subsets based on different parameters without introducing bias. The experiment included a total of 135 design points (i.e., 135 unique combinations of levels of the parameters), with 100 replications run for each design point; yielding a total of 13,500 simulation runs.

Results and Analysis

The central question guiding the analysis of results was, "Which ADD configuration produces the most favorable results?" WRENCH can produce outputs toward a wide range of performance metrics. In this analysis we focus on two: the number of intruders (a primary metric of mission success for compound defense), and the amount of escalation in the average hostility level of the crowd (a potential consequence of the use of force that has implications on the mission going forward). We also discuss how examining the influence of other experimental parameters, particularly the characteristics of the population, can provide more insight into the effectiveness of different ADD configurations. In this section, all averages were taken across all replications with the stated data subset, unless otherwise specified. And in the bar charts, the "whiskers" indicate 95% confidence intervals.

Number of Intruders

As can be seen in Figure 2, the average number of intruders was quite high under each IFC option. Considering that no weapons were in use that could significantly injure anyone or cause death, this makes sense. The results also show that there was not a great deal of variation in the number of intruders under each IFC option, although a one-way ANOVA test confirmed significance at the $p < 000.1$ level. The ADD-1 configuration resulted in the lowest average number of intruders (24.8), ADD-2 resulted in the most (29.6), and the other three options resulted in a moderate number of intruders, with a small amount of variation among them (27.8, 27.4, 28.3, respectively).

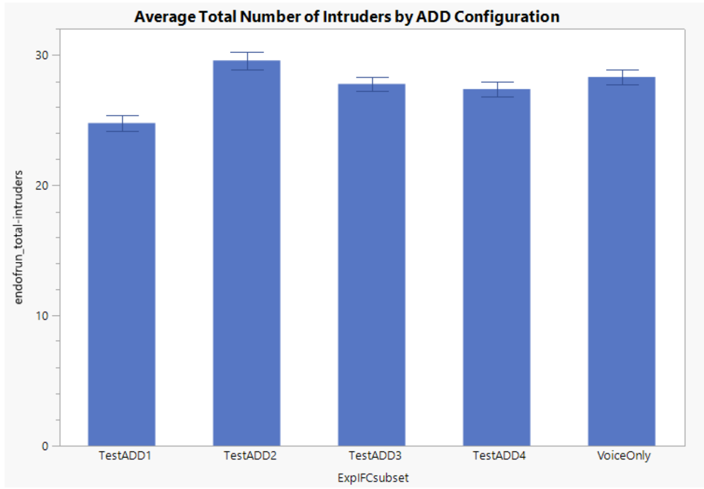


Figure 2. Average Number of Intruders by ADD Configuration (Treece, 2024)

Crowd Hostility

In WRENCH, hostility is measured on a [0,6] scale, where 0 is completely compliant and 6 is deadly hostility. And, while a person doesn't automatically act on their hostility, it is a significant driver of behavior; it can cause them to begin protesting, or to aggressively move toward the SF while protesting, or even decide to attack (attempt to invade the compound), which also has ripple effects through groups and the crowd.

Figure 3 shows the average hostility level of the people (averaged across the people within each replication, then averaged across the replications). As can be seen, ADD-1 and ADD-2 greatly escalated the average hostility of the people, while the other three options minimally escalated hostilities. And when comparing Figures 2 and 3, there appears to be somewhat of a trade-off between intruders and average hostility. Notably, while ADD-1 is the best according to the intruders metric, it is actually the worst on the hostility metric. This shows the importance of using a simulation model that can provide outputs on multiple metrics of interest, and of considering performance on all of those metrics when determining what IFC configuration option is "best."

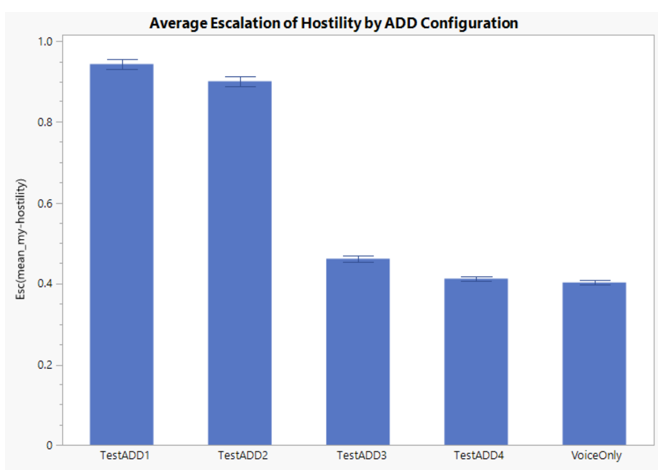


Figure 3. Average Escalation of Hostility by ADD Configuration (Treece, 2024)

Why Population Matters (how above "answers" regarding "best" ADDs may differ based on population type)

When aggregating data, such as was done in the averaging of the results across all replications as seen in the above analysis, important information can be lost (Aros & McDonald, 2023c). A primary benefit of conducting a full-factorial experiment is that it is possible to split the data into subsets based on different parameter values without introducing bias, allowing for disaggregation of the data for further analysis. In preliminary exploration of results, it became clear that the outcomes differed greatly based on population type, with the largest differences being between the Market population as compared to each Protest population. So here we discuss the results separated by population type, looking at the Market population results separately from the aggregate of both Protest populations on the two metrics of interest.

Figure 4 shows these results for the average number of intruders metric. Not surprisingly, the average number of intruders in Protest populations is much higher than for the Market population, across all IFC options. We also see that ADD-1 did achieve the best results, (i.e., lowest intruders) under both the Market population (5.6) and the Protest populations (34.4), consistent with the fully aggregated results discussed above. However, upon closer examination, some differences become clear. For example, we see that, for the Protest populations, ADD-2 gives the worst outcome (40.4 intruders), whereas for the Market population the ADD-2 gives the second-best outcome (8.0 intruders).

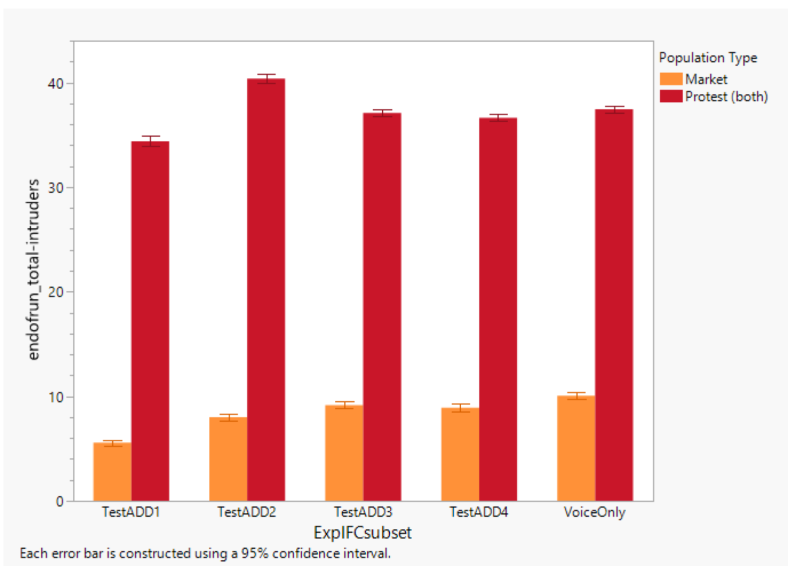


Figure 4. Average Number of Intruders by ADD Configuration, Split by Population Type

Figure 5 provides the hostility escalation results, split by population. Here we can see that, for the Protest populations, the ADD-1 and ADD-2 options result in an average of 138% higher hostility escalations than result from the other three IFC options, averaged. But for the Market population, ADD-1 and ADD-2, averaged, only result in only 56% more hostility escalation than the other three IFC options, averaged.

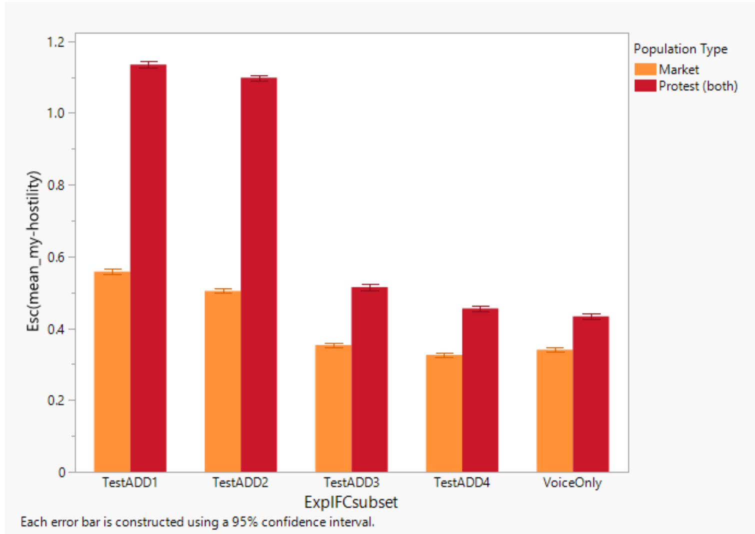


Figure 5. Average Escalation of Hostility by ADD Configuration, Split by Population Type

Overall, this examination of results over different populations, and across two different performance metrics, shows that what had appeared to be the clear dominance of ADD-1, based on the intruders performance metric over the fully aggregated dataset, is actually not so clear. The very high escalation of hostility caused by the ADD-1 should also be considered. Also, the fact that the escalation of hostilities caused by ADD-1 and ADD-2 in the Protest populations is so high, relative to the other IFC options, may indicate to decision-makers that it may be worth considering one of the options that was not optimal on the intruders metric. A full analysis of the results that examines difference in results across other parameters, and across other important metrics such as the escalation in the number of protesters and attackers, would further highlight important nuances in the trade-offs between ADD configuration options. While the new insights from this second-level analysis by population type are few, Aros & McDonald (2023b) demonstrate to what extent the combination of parameter values that produce the most desirable outcomes can be quite different for different populations.

Discussion

Here we discuss how WRENCH can support the acquisition process in different phases. We also discuss important limitations of using simulation modeling in the acquisition process.

How WRENCH Can Inform the Acquisition Process

In the AoA activities of the MSA phase, and in various TM&RR phase activities, WRENCH could be used to simulate the use of different types of theoretical NLW systems, capabilities, or proposed weapons designs, enabling the analysis of the relative effectiveness of different options. This analysis could also provide insights into what factors, whether of the weapons themselves or of the situations or methods of use, most contribute to different effectiveness outcomes. These insights could then be used not only in the selection among alternatives, but also to inform the exploration of theoretical alternatives not yet considered.

During the EMD life-cycle phase WRENCH could be used in DT&E activities to test different design specifications or characteristics of the NLW system to see which is most effective, similar to what was presented in this paper. In addition, these design options could be tested across broad ranges of potential use scenarios to see if the relative effectiveness across the design options differs significantly in different scenarios. And as a part of OT&E activities, WRENCH could be used to test different ways in which the new technology could be used in conjunction with existing non-lethal and lethal capabilities. This can highlight any synergies of challenges in integrating the new technology into the SF force continuum, but can also be used to aid in the development of guidance for how to best use the new technology.

During the P&D phase, once the technology design has been finalized and is ready to be produced, WRENCH can be used to explore different deployment strategies for the new technology such as which force members should be issued the technology. Also, once the technology has been finalized, further exploration of the most effective use strategies can be explored as well.

Limitations of Simulation Modeling

When using any type of simulation, it is important to keep in mind that a simulation will never fully replicate reality, even "live" simulations. Therefore it is important to consider the purpose of the simulation and determine if it is "good enough" for that stated purpose, being careful not to use simulation to inform decision-making that it was not designed to support. For WRENCH, the stated purpose is to evaluate the relative differences in the metrics of interest across realistic scenarios; it was not designed to predict actual outcomes for specific situations. Extensive efforts have gone into the verification and validation of WRENCH, though, and these efforts are ongoing.

Conclusions

This paper discussed the use of computer simulation modeling to assess, test, and evaluate NLW systems. Specifically, we demonstrate how the WRENCH simulation model can be used to explore the effectiveness of different NLW configurations on the achievement of mission objectives in a compound defense civil security scenario. We discussed the design of the experiment, analyzed the results, and provided recommendations. Though our experiment was based on a hypothetical ADS, our findings and discussion indicate that WRENCH could be used throughout the acquisition life cycle for the development of NLW, especially during the MSA AoA activities, and the TM&RR, EMD, and P&D phases.

Our work also serves as a demonstration of how the DoD can leverage simulation capabilities throughout the defense acquisition life cycle for the assessment, testing, and evaluation of products. This would require the selection or development of simulation models suitable for the specific weapons and testing environments, as WRENCH is suitable for exploring NLW effectiveness. The use of computer simulation throughout the acquisition life cycle can be an innovative way for transitioning emerging technologies from research and development to defense acquisition programs of record and thus help the program successfully cross the valley of death.

The experiment and analysis presented in this paper give just the smallest glimpse into the capabilities of WRENCH for testing NLW during the acquisition life cycle. WRENCH can generate outputs on a wide variety of metrics of interest, and can be used to test the effectiveness under a wide variety of conditions and weapon configuration differences, providing a wealth of data for analyses that can shed light on what factors contribute most to the effectiveness of the weapons toward different metrics of interest. WRENCH can also easily be updated to model any new NLW type or new NLW configuration for testing, whether existing or in the planning stages of development. WRENCH currently supports simulation of compound security missions, but could also be extended to modeling other civil security situations such as border patrol and humanitarian aid distribution.

We have a number of ongoing research avenues pertaining to the use of WRENCH. Most immediately, we are continuing to analyze the dataset from the experiment presented in this paper in order to provide more insights into the combinations of different parameters, and which of the specific differences between ADD configurations most affect the results.

We also have multiple efforts underway in the ongoing improvement, verification, and validation of WRENCH. In addition, a future avenue of work just begun is to explore how the simulation advancements made with WRENCH could be leveraged and adapted to construct a similar type of simulation for modeling the use of IFCs in the maritime gray zone. WRENCH is fully DoD owned and developed, leaving the door open for limitless further development and adaptation.

References

Afara, J., Ajila, V., Macdonell, H., & Dobias, P. (2024). Use of agent-based modeling to model intermediate force capabilities in (counter) mobility crowd scenarios. *Journal of Defense Modeling and Simulation*, *21*(4), 395–404.

Aros, S. K., Baylouny, A. M., Gibbons, D. E., & McDonald, M. (2021) Toward better management of potentially hostile crowds. In *2021 Winter Simulation Conference (WSC)*. IEEE.

Aros, S. K., & McDonald, M. (2023a). *Simulating civil security activities in stability operations*. Presentation for the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Orlando, FL.

Aros, S. K., & McDonald M. (2023b). Simulating civil security activities in stability operations. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. Orlando, FL.

Aros, S. K., & McDonald, M. L. (2023c). Squashing bugs and improving design: Using data farming to support verification and validation of military agent-based simulations. In *2023 Winter Simulation Conference (WSC)* (pp. 106–117). IEEE.

Buch, B., & Mitchell, K. (2013). The active denial system: obstacles and promise. In *Convention on Certain Conventional Weapons*.

Burgei, W. A., Foley, E. E., & McKim, S. M. (2015). Developing non-lethal weapons: The human effects characterization process. *Defense AT&L*.

DoD. (2022, June 8). *Operation of the adaptive acquisition framework* (DoDI 5000.02). Office of the Under Secretary of Defense for Acquisition and Sustainment.

GAO. (2024).

Gray, S. P. (2017). *Agent-based simulation to support the effectiveness, procurement, and employment of non-lethal weapon systems*. Naval Postgraduate School.

Grimes, J. D. (2005). *Modeling sound as a non-lethal weapon in the COMBATXXI simulation model*. Naval Postgraduate School.

Grocholski, K. R., Savitz, S., Wong, J. P., Litterer, S., Khan, R., & Cooper, M. (2022). *How to effectively assess the impact of non-lethal weapons as intermediate force capabilities*. RAND.

LeVine, S. D. (2009). *The active denial system*. Center for Technology and National Security Policy, National Defense University.

Mezzacappa, E. (2014). Effectiveness testing of non-lethal weapons. *Journal of Defense Modeling and Simulation*, *11*(2), 91–101.

Peters, S. D., Leiden, K. J., Ehrhart, L. S., Bigbee, A. J., & Goodrich, C. G. (1998). Modeling human response to non-lethal weapons in distributed simulation environments. In *Proceedings of the Seventh Conference on Computer Generated Forces and Behavioural Representation*. Orlando, FL.

Treece, D. E. (2024). *Active denial technology and effects on crowd behavior*. Naval Postgraduate School.

Vila, R. (2010). *Modeling and simulation synergies in acquisition, test and evaluation, and training*. International Test and Evaluation Association (ITEA) Live-Virtual-Constructive Conference 2010.

Wang, H., Burgei, W. A., & Zhou, H. (2020). A concise model and analysis for heat-induced withdrawal reflex caused by millimeter wave radiation. *American Journal of Operations Research*, *10*(2), 31–81.

Wirthlin, J. R., Houston, D. X., & Madachy, R. J. (2011, May). Defense acquisition system simulation studies. In *Proceedings of the 2011 International Conference on Software and Systems Process* (pp. 189–193).

Revolutionizing Marine Corps Training: The Marine Corps Reconfigurable Consolidated Driver Simulator (MCRCDS) Initiative

Dr. Leili Green—Human Systems Integration & Instructional Systems Specialist & Mobile Training System Training Lead, Global Combat Support System - Marine Corps (GCSS-MC). Dr. Green brings 25 years of experience in training and simulation across government and private sectors, including medicine, higher education, energy, the DoJ, and the DoD. She specializes in human systems integration and instructional design, developing and implementing training systems for the USMC, focusing on simulator acquisition. [Leili.green@usmc.mil]

Abstract

The Marine Corps has made significant strides in recent years by acquiring modernized driver simulators to strengthen training effectiveness and operational readiness. However, the long-standing platform-centric approach—common across all military services—limits the ability to fully leverage emerging modular hardware, open architecture systems, and advanced terrain software that could support multi-platform use.

This paper introduces the Marine Corps Reconfigurable Consolidated Driver Simulator (MCRCDS) initiative, developed to address the fragmentation, inefficiencies, and high costs caused by over 48 standalone simulators across various commands. MCRCDS offers a reconfigurable, consolidated solution that incorporates Artificial Intelligence (AI), Machine Learning (ML), and the Explore, Experiment, and Excel (EEE) learning principle to provide immersive, adaptive, and personalized training aligned with the Commandant's vision for 21st-century readiness.

By integrating Live, Virtual, and Constructive (LVC) environments and standardizing data collection, MCRCDS supports broader Department of Defense goals to improve readiness predictions and identify training gaps. The initiative's phased implementation and cost analysis underscore its potential to significantly reduce administrative burden, enhance interoperability, and improve training outcomes.

This research highlights MCRCDS as a transformative solution for modern driver training—moving the Marine Corps toward a more efficient, scalable, and future-ready simulation capability.

Introduction

The United States Marine Corps has taken significant strides in modernizing the simulators/training systems for the land systems, particularly driver and craw simulators for their tactical vehicles. Moreover, they continue incorporating the latest technologies into their future drive estimator acquisitions.

The platform-centric approach in training systems is not unique to the Marine Corps or a specific service or particular capability tactile vehicle type.

The Commandant of the Marine Corps continues to emphasize technological advancement by adopting the latest technologies to maintain momentum and through collaboration with industry, academia, joint forces, and allies. The Marine Corps Systems Command DC SEAL initiative of the MCRCDS study is aligned with reducing duplication of efforts, maintaining technological superiority momentum, and reducing noncombatant vehicle fatal mishaps.

The Marine Corps is at a critical juncture, needing to modernize its training systems to meet the demands of 21st-century warfare. Once effective, the existing driver training programs are now spread across different Formal Learning Centers (FLCs) and commands, as shown in

Table 1. Due to the complexity and cost associated with changes to make the existing driver simulators interoperable, modernizing them is not feasible. This study will show the existing USMC driver simulator landscape and propose a new solution.

Problem Statement

The existing Marine Corps driver training systems are platform-centric and fragmented across different vehicle types, programs, and commands, leading to high operational costs, interoperability issues, and limited scalability. This fragmented approach undermines training effectiveness without significantly reducing non-combat driving mishaps.

Current Marine Corps Tactical Vehicle Driver Simulators Distribution

This section details the current distribution and future growth of Marine Corps tactical vehicle driver simulators. It highlights the fragmented nature of existing systems across various vehicle platforms.

As shown in Table 1, the Marine Corps driver training systems are currently dispersed across more than 48 standalone simulators.

Table 1: Breakout of 48 Operator Driver Simulators

Location	Cab Number	Instructor Operation Station(s)- IOS	Mobile Trailer
MCAS Iwakuni	2	1	X (Dual)
MCB 29 Palms	4	2	X (two dual)
MCB Camp Pendleton-California	6	4	X (two dual and two single)
MCB Camp Lejeune, North Carolina	5	3	X (two dual, one single)
MCB Okinawa	5	2	
MCB Hawaii	2	1	X (dual)
Fort Leonard Wood, Missouri	24	6	
Totals	48	19	

Please note that four Light Armored Vehicle Driver Trainers (supplied by Wegmann USA) are currently at the School of Infantry West at Camp Pendleton (SOI-W). The number of existing driver simulators for recent additions to specific programs has not been included and is in addition to the depicted number here.

Overview of Sources

This study builds upon the foundational concepts outlined in the Commandant's guidance, emphasizing technology integration. Insights collected from engaging with stakeholders across the Marine Corps and industry show the need for modernizing driver training systems.

The same concerns and solutions were echoed in the book *Kill Chain* by Christian Brose (2020), which emphasizes the need for a leap to 21st-century technology and a shift from platform-centric to network-centric approaches in defense acquisitions. In addition, Mark A. Miley and Eric Schmidt's paper published on August 5, 2024, in Foreign Affairs Magazine titled

“America Isn’t Ready for the Next Wars of the Future” raised the same concerns about the platform-centric approach. The MCO 3550.14 listed on the resources page and multiple documents related to PM TRASYS were also reviewed (U.S. Marine Corps, 2021). The review included the trade-off study conducted for the common Marine Corps Driver Trainer and articles on the ACV driver and Operator driver simulators published on Inside Defense.com and Breaking Defense.com.

Mark A. Miley and Eric Schmidt’s paper, "America Is Not Ready for the Next Wars of the Future," published on August 5, 2024, in Foreign Affairs Magazine, voiced the same concerns about the platform-centric approach. The MCO 3550.14 listed on the resources page and multiple documents related to MC- Program Office Training Systems PM TRASYS were also reviewed regarding the tradeoff study conducted for the standard Marine Corps Driver Trainer and articles on the ACV driver and Operator driver simulators published on Inside Defense.com and Breaking Defense.com.

Study Approach

This study approach adopted here is like the aviation practice of a 360-degree clearing turn, where every aspect of the area is visually assessed prior to proceeding. This study has done the same with the existing driver simulators and the recommendation for change.

The interconnected elements of the Marine Corps’ current operator driver training systems were reviewed from many angles and phases of acquisition to sustainment. The associated current and future learning principles, policies, and previous studies for a standard driver simulator platform conducted by MCSC PM TRASYS 2023 were examined. A collaborative effort of engaging with stakeholders and subject matter experts provided information reflected throughout the paper.

Findings

How Did We Get Here & What Are the Symptoms?

As previously mentioned, the existing landscape of the Marine Corps driver training systems is dispersed due to a historically siloed approach to system capability-based acquisition. Each automotive vehicle program across the command and entities independently acquires its training systems, leading to a proliferation of standalone simulators (see Table 1). The capability-centric acquisition requirement and approach have resulted in redundancies and increased operation and sustainment costs. The symptoms of this fragmentation are evident in high costs, limited scalability, and the inability to standardize data across systems. The existing data extracted vehicle variants lack format uniformity due to proprietary issues and dispersed software licenses.

Additionally, non-combat driving mishaps have not decreased. Ground vehicle mishaps were the leading cause of deaths and injuries in the line of duty in the U.S. military between 2010 and 2021, according to Government Accountability Office Report 21-361. Most ground vehicle mishaps are caused by driver errors in judgment and deficient skill rather than external factors.

The following key factors that contributed to led us this point:

- Siloed Training Systems: Training programs are tailored to specific vehicle types, resulting in duplication of simulators, software, and maintenance. Per MCO 3550.14, funding for acquiring driver simulators is currently allocated to each program office within the PEO LS and other entities for Standard and Nonstandard Training Systems.

MCO 3550.14 defines standard and nonstandard training systems:

"A Standard Training System is a training solution developed and acquired for use with a specific system (e.g., weapons platform, vehicles), family of systems, or item of equipment (including subassemblies and components). Standard Training Systems may be standalone, embedded, or appended.

"Nonstandard Training Systems is a training solution developed and/or acquired independent of, and not directly associated with, a specific weapon system or other item of equipment. Nonstandard training systems may support general military training, system-specific, and non-system specific training requirements."

- High Costs: Each vehicle type requires a separate training system, which increases the costs of hardware, software, operation, and maintenance.
- Interoperability Issues: The presence of proprietary standards for each simulator has created a fragmented landscape, hindering interoperability, data exchange and standardization, and functionalities. In addition to the ascending operation costs of standalone systems, there is a lack of interoperability collaborative efforts, as aimed to be achieved by the TECOM Live Virtual Constructive (LVC) effort.
- Scalability Limitations: Current simulators cannot integrate new functionalities and collect biometric and training incidents in a standardized data format.
- Simulation System Design Flaws: Existing systems lack modular hardware and open-architecture software, which leads to inefficiencies in training module integration and data analysis.

The identified inefficiencies impact costs and contribute to the need for an enterprise driver operator simulator critical to maintaining combat readiness. However, these inefficiencies also present an opportunity to rethink and reimagine the future of Marine Corps individual and convoy driver training. The MCRCDS concept offers a pathway to provide a solution to address inefficiencies and set a new standard for driver training excellence across the Marine Corps.

Discussion

As pointed out in the above sections, the existing USMC driver simulator capability-centric acquisition approach has supported this study.

One measure to address data standardization issues is leveraging the computing power available to collect timely reports and objective analysis from training systems while also being able to extract valid reports that can be used as input to force readiness predictable models. Reconfigurability and consolidated platform for driver operator simulators must include comprehensive biometric data and real-time training incident scenarios using AI and ML. The need to augment the existing Crawl Walk Run (CWR) with Kolb's Experiential Learning Theory Explore, Experiment, and Excel (EEE) learning principle will support the integration of AI in future driver simulation.

Learning Principles and Data Standardization

The current system design acquisition is based on the CWR learning principle. The current training simulators require manually created training evaluation and lack standardized objective evaluation based on human and machine teaming concepts with embedded scoring algorithms to assess performance. The current training delivery method places a higher demand on the instructor, and after-action reviews must be held to identify student mistakes if they are noticed retrospectively. Currently, the training system acquisition and sustainment model omits the requirement for a standard open architecture driver simulator software or AI and ML technology integration, necessitating a paradigm shift. To ensure a successful leap to 21st-century training, augmenting the existing CWR learning principle with the EEE learning principle grounded in Kolb’s for future AI-enabled driver training simulators is crucial. There are no known development initiatives to address inefficiencies leveraging current computing power.

The power of AI and ML is reshaping traditional training methods, such as the CWR principle. Integrating AI and ML augmented by the EEE principle promotes an ongoing learning cycle that encourages trainees to explore new concepts, experiment with them, and excel by refining their skills to meet the demands of continuous learning. This adaptability required in today’s dynamic technological environment is presented in the Marine Corps Project Tripoli, and the Army’s Synthetic Training Environment (STE) projects are discussed . The Marine Corps Project Tripoli and the Army’s STE, showcased on Marine Corps Times and Army.mil, demonstrate how LVC training elements enhanced by AI and ML offer a personalized and immersive training experience. The training starts with basic operations and progresses to complex and adaptive scenarios. The adaptive LVC training environment walks Marines through the basics and urges Marines or soldiers to experiment with different strategies and learn from their experiences.

The combination of an AI-empowered driver simulator augmented with CWR and enhanced with the EEE learning principles ensures that trainees are proficient in fundamental skills and equipped with critical thinking and adaptability, which are essential for modern warfare.

AI involves developing computer systems that can perform tasks requiring human intelligence, such as learning, reasoning, problem-solving, perception, and understanding natural language. Machine Learning is a subset of AI that automatically enables systems to learn and improve from experience. Standardization is the process of establishing and applying consistent protocols, criteria, and specifications across systems. Finally, biometric data refers to individuals’ unique physical or behavioral characteristics. Integrating AI, ML, biometric data, and data standardization in driver simulators can generate data to predicate force readiness and identify training gaps. This data helps plan for the desired level in enhancing precise and relevant training scenarios of individualistic and standardized collective training.

The Chief Engineer Role

The Chief Engineer (CHENG) plays a crucial role in this initiative by overseeing these technologies’ integration standards to ensure the system’s scalability and interoperability while addressing the technical challenges associated with the transition.

Risks and Risks Mitigation

The transition to the MCRCDS carries certain risks. Incorporating AI and data standardization into existing systems presents challenges, including potential data security concerns and the complexity of operating high technological system costs. However, these risks can be mitigated with proper planning, continuous testing, and phased implementation. While the proposed unified training system offers numerous benefits, it has potential risks. Understanding and mitigating these risks is crucial for successful implementation and operation. The primary risks associated with the new system include:

- Technical Complexity: Integrating advanced technologies such as AI and biometric data collection introduces technical complexity that may lead to unforeseen challenges during development and deployment. These technical challenges could negatively impact cost, schedule, and performance.
- Data Security and Privacy: The new system will handle sensitive biometric and performance data, raising concerns about data security and privacy. Ensuring robust cybersecurity measures and compliance with data protection regulations will be critical.

- Cost Overruns and Reduction Measures: The new system's complexity and scale might lead to higher-than-anticipated costs. Budget overruns could occur due to unexpected technical challenges, extended development timelines, or additional unknown resource requirements. Although the initial estimate proposed in this study may seem lower than the average cost in the market, the use of Model-Based Systems Engineering (MBSE), digital documentation, and Commercial Off-the-Shelf (COTS) hardware solutions can help reduce the cost.
- Risk Mitigation: Risk mitigation aims to adopt robust project management, including conducting detailed cost analysis, enhanced cybersecurity measures, and the system's interoperability and scalability to reduce operation and maintenance costs. By proposing a consolidated system, the assumption is to reduce the number of driver simulator systems. The improvements in the reconfigurability and efficiency of the new systems suggest that the decrease in driver simulator systems will not significantly affect throughput or quality of training. A phased fielding and transition approach must mitigate any negative impact on force readiness. By proactively addressing these risks and implementing appropriate mitigations, the potential challenges associated with the new unified training system can be effectively managed, ensuring a successful transition and long-term operational success.

Phased Approach and Cost Analysis

A programmatic phased approach is recommended to ensure a structured and efficient transition to the MCRCDS. This approach allows for stakeholder engagement, phased development, and iterative refinement, ensuring the final system meets all operational requirements and technical standards. The initial study phase focuses on gathering detailed information on the current state of the Marine Corps driver simulators, identifying gaps and challenges, and exploring the potential benefits of transitioning to a consolidated, reconfigurable system.

Study Limitations

Studies are inherently limited by their preliminary nature; with buy-in from stakeholders, the findings must lead to viable solutions. Due to the complexity of the fact-finding process, the large number of existing assets, and time constraints, this study primarily serves as a guide to inform decision-makers.

Benefits to the Marine Corps

The MCRCDS is not just a solution to the current standalone driver simulator problems facing the Marine Corps; it is a strategic investment in the future. By standardizing data, integrating new technologies, and transitioning to a network-based enterprise solution, the Marine Corps will modernize its training standard and nonstandard training systems fleet, reduce costs, and improve readiness. This approach aligns with the Commandant's vision for 21st-century warfare. By embracing new learning principles and technologies and adopting a network-centric approach, the Marine Corps will be better equipped to meet the challenges of the 21st century. The lessons learned from the MCRCDS can be applied to other training systems, such as gunnery and combat crew training systems, ensuring that the Marine Corps remains at the cutting edge of military readiness.

Conclusion and Next Steps

The current landscape of driver simulators across services, including Marine Corps driver training systems, is characterized by a fragmented and inefficient structure that contributes to duplicated efforts, inflated costs, and diminished system performance. The challenges are multifaceted, extending beyond the technical difficulties of software updates, hardware replacements, and sustainment issues. They include a burdensome logistics footprint that directly impacts force readiness. This status quo is unsustainable, and a cohesive, scalable, and future-proof solution is critical.

The proposed MCRCDS is designed to address these challenges head-on. By net-centric consolidated driver-operator simulators across multiple vehicle types into a unified, adaptable system, the MCRCDS represents a significant advancement in operational efficiency. The proposed solution streamlines infrastructure reduces maintenance and operational expenses, and lays the foundation for a more responsive and agile training capability that can evolve with the Marine Corps' changing needs.

One of the MCRCDS's key advantages is its modular design, which ensures the system can quickly integrate new vehicle types and adapt to emerging technological advancements. By leveraging AI, ML, and standardized data protocols, the system can dynamically adjust training scenarios and provide on-demand, tailored experiences that meet mission-specific needs. This adaptability enhances training effectiveness and ensures scalability for future requirements, making the MCRCDS a rapid response solution capable of evolving with modern warfare demands.

Investing in the MCRCDS is a strategic decision that aligns with the Commandant's vision for 21st-century warfare. By adopting a net system-centric approach and incorporating advanced learning principles, the Marine Corps will significantly enhance its operational readiness and efficiency, ensuring its training systems remain at the cutting edge of military capability. In addition to technological advancements in the future generation of driver simulators, reducing fatal non-combat driving training mishaps is essential.

Strategic Implications

The adoption of the MCRCDS is far more than a technical enhancement; it represents a strategic investment in the Marine Corps' future operational capabilities. By transitioning from platform dependency to a net system consolidated solution and standardizing driver-operator training systems, the Marine Corps will be better positioned to meet the challenges of modern warfare. The efficiencies gained from this system will reduce costs and ensure that the Marine Corps can maintain a high state of readiness across all vehicle platforms.

Moreover, the principles and lessons learned from implementing the MCRCDS can be extended to other critical training systems, such as gunnery and combat crew training. This holistic approach will ensure that the Marine Corps remains at the forefront of military readiness, capable of adapting to the evolving demands of 21st-century warfare.

Sources

AI in supply chain: Challenges, benefits, & use cases. https://intellias.com

Bohemia Interactive Simulations. https://bisimulations.com

Brose, C. (2020). *The kill chain: Defending America in the future of high-tech warfare*. Hachette Books.

Brown, T., & Wang, L. (2023). Cost-benefit analysis of consolidated training systems. *Journal of Defense Economics*, *31*(1), 55–70.

Cole Engineering Services, Inc. https://coleenginerring.com

Defense Advanced Research Projects Agency. (2023). *AI and machine learning applications in military training*. DARPA Technical Report.

Department of Defense. (2022). *Virtual wargaming: Enhancing military preparedness through simulation*. U.S. Government Printing Office.

GAO. (2021). Report 21-361.

Johnson, A., & Patel, R. (2023). Advancements in biometric feedback for training systems. *Defense Innovations Journal*, *29*(2), 22–35.

Lockheed Martin. https://www.lockheedmartin.com

MAK Technologies. https://www.mak.com

Marine Corps fielding ACV driver simulators, preparing for wider vehicle deployment. (2024) https://insidedefense.com/

Marine Corps Systems Command. (2021). *Operational requirements for driving simulators*.

Marines eyeing driver simulator for ACV training. (2023). https://breakingdefense.com/2023/06/marines-eyeing-driver-simulator-for-acv-training/

MCSC Program Office Training Systems (PM TRASYSSYS).

Miley, M. A., & Schmidt, E. (2024). American isn't ready for the next wars of the future. *Foreign Affairs*.

Office of Naval Research. (2022). *Future trends in military simulation and training*. ONR Publications.

Smart transportation logistics: Achieving supply chain efficiency. https://link.springer.com

Smith, J. (2023). Integrating AI into military training simulators: Benefits and challenges. *Military Technology Review*, *45*(3), 34–47.

U.S. Marine Corps. (2021). *MCO 3550.14 Policies and procedures for synthetic ground training systems*. https://www.marines.mil/News/Publications/MCPEL/Electronic-Library-Display/Article/2668304/mco-355014/

U.S. Marine Corps Training Command. (2022). *Annual report on driving simulator programs*.

Finding Opportunities in the Adaptive Acquisition Framework

Robert Paul Lewis—Systems Engineer and Architect with 10+ years of relevant experience, including conceptual design, needs identification, mission engineering, requirements definition, and validation in DoD, Army, and Air Force contexts. Architected, designed, and developed a hypersonic weapon system for the B-52H platform as part of a Systems Design Agent contract with the Air Force Research Laboratory. Extensive experience in applied Model-Based Systems Engineering principles and languages, such as SysML, to create and manage system models, architectures, and interfaces. Performed functional decomposition and allocation of system functions and behaviors to subsystems and components, ensuring traceability and consistency with system requirements. Established expertise in conducting trade studies, risk assessments, system performance, and MOSA architectural evaluations of complex weapon systems at the enterprise, FoS/product line, system, subsystem, and component levels. [rplewis@att.net]

Joseph Carnes—Acquisitions SME, Sariel, LLC [joelinux@gmail.com]

Jack Thompson—President, JTE Systems, LLC. [jt@jtesys.com]

Abstract

For over 4 decades, the Department of Defense (DoD) has pursued Major Capability Acquisition (MCA) reforms to counter rising threats, yet programs like the F-35 and Zumwalt-class destroyers suffer persistent cost overruns, delays, and performance shortfalls. This study analyzes DoD policies, Government Accountability Office and RAND Corporation critiques, and external scholarship to reveal why modularity goals, like the Modular Open Systems Approach, falter despite aims for innovative, adaptable systems with strong lifecycle outcomes. Three flaws persist: requirements obscuring utility across cost, schedule, performance, and lifecycle; centralized contractor structures embedding complexity; and contract scales eroding DoD control. With $183 billion in overruns across 36 programs (GAO, 2023), MCA's misalignment—contractors favoring profit incentive over warfighter value—demands change. Historical successes inspire a solution: World War II's (WWII's) 18,000 firms delivered 297,000 aircraft, showcasing modularity and adaptability. To address MCA's centralized failures, a distributed acquisition model is proposed, fractionating systems into small teams of up to 150 members. This approach fosters competition, simplicity, and responsiveness, leveraging organizational theory and analytical tools to meet DoD goals. While implementation awaits further study, this shift promises significant savings and operational agility, urging acquisitions professionals to move beyond reform tweaks and embrace a proven alternative rooted in history.

Introduction

For over 4 decades, the Department of Defense (DoD) has pursued Major Capability Acquisition (MCA) reforms to deliver modular, innovative systems with positive lifecycle outcomes, adaptable to an evolving operational environment filled with rising peer and near-peer threats. Despite these efforts, MCA programs consistently falter, undermining the DoD's vision as articulated in foundational policies like the National Defense Strategy (DoD, 2022b). This study defines persistent flaws thwarting MCA's goals and proposes a distributed acquisition model as a transformative solution. This study asks, what are the DoD's stated intents and challenges in MCA? How do immediate stakeholders, such as the Government Accountability Office (GAO) and RAND Corporation, assess these efforts? What solutions do external experts propose? Through a comprehensive review of DoD policies, stakeholder critiques, and external scholarship, this paper identifies three structural issues: First, requirements miscommunicate military utility across cost, schedule, performance, and lifecycle. Second, centralized contractor organizational structures embed complexity into the solution. Finally, the increasingly large contract scales erode DoD control. This demonstrates the need for a conceptual shift to realign MCA with DoD objectives, while reserving implementation details for a follow-on study.

The DoD's reform journey spans multiple initiatives, from the Goldwater–Nichols Act of 1986, which centralized authority to streamline processes, to the Weapon Systems Acquisition Reform Act (WSARA) of 2009, Better Buying Power (BBP) initiatives, and the Adaptive Acquisition Framework (AAF) of 2019, each targeting cost overruns and delays (DoD, 2023; GAO, 2012b). The Modular Open Systems Approach (MOSA), mandated in 2017, seeks modularity to enhance competition and adaptability (DoD, 2022b). Yet outcomes remain dire. The GAO reports, "In 2023, MCA programs accumulated $183 billion in cost overruns and average delays of two years across 36 programs," with only 14 of 20 programs partially adopting MOSA (GAO, 2023, p. 1). This disconnect between intent and execution signals deeper, structural failures.

Two programs exemplify these challenges. The F-35 Joint Strike Fighter, developed by Lockheed Martin, has faced significant delays and cost growth due to requirements granting priority to stealth and multirole capabilities over affordability and sustainment, straining budgets and operational timelines (GAO, 2015). Similarly, the Future Combat Systems (FCS), canceled in 2009 after investing $18 billion, aimed for a networked system but collapsed under technical complexity and unclear lifecycle goals (Pernin et al., 2012). These cases highlight a pattern: despite reform efforts, MCA struggles to deliver modular, adaptable systems, with costs and delays eroding warfighter readiness.

Theoretical frameworks illuminate these issues. Conway's Law posits that system designs mirror organizational structures, suggesting that centralized contracting organizations produce complex, integrated systems ill-suited for modularity (Conway, 1968; MacCormack et al., 2012). A centralized contractor refers to a single, typically lead system integrator or prime contractor that assumes primary responsibility for designing, developing, integrating, and delivering an entire complex weapon system or major program within the DoD acquisition process. This entity consolidates control over most or all subsystems, often subcontracting components but retaining overarching authority under a monolithic contract structure. Centralized contractors dominate the acquisition process through their extensive resources, proprietary technologies, and entrenched relationships with the DoD, exemplified by firms like Lockheed Martin (e.g., F-35), Boeing, SAIC (e.g., FCS), or Northrop Grumman.

Brooks (1995) reinforces this, noting that large teams exacerbate delays and complexity, a reality MCA reflects. Principal–Agent Theory reveals a further misalignment: contractors prioritize their profit motive over warfighter utility across cost, schedule, and lifecycle phases, as seen in FCS's integrator-driven focus (Pernin et al., 2012). McChrystal (2015) critiques rigid hierarchies as ill-equipped for dynamic threats, underscoring MCA's structural rigidity.

The stakes are high. Emerging threats from adversaries demand systems that innovate and adapt, yet MCA's centralized framework—exacerbated by a shrinking Defense Industrial Base (DIB) and overwhelming contract scales—locks the DoD into a cycle of inefficiency. Historical successes, such as distributed acquisition during WWII, contrast sharply with this reality, suggesting a path forward. External scholarship supports this, with analytical tools like Value-Driven Design (VDD) and Multi-Attribute Utility Theory (MAUT) offering ways to optimize utility and reduce complexity (Abbas, 2018; Collopy, 2007). Organizational insights from small-team successes further bolster the case for change (Brooks, 2010; McChrystal, 2015).

This study's proposed distributed acquisition approach, which emphasizes decentralized structures to enhance modularity and adaptability, aligns with emerging legislative efforts to address MCA's systemic issues. Notably, the Fostering Reform and Government Efficiency in Defense Act (FORGE Act), introduced in December 2024, seeks to streamline DoD acquisition by reducing bureaucratic barriers, prioritizing commercial contracting, and fostering competition to diversify the DIB (Wicker, 2024). By advocating for agile, distributed approaches over centralized complexity, this paper's framework complements the FORGE Act's vision, offering a conceptual foundation to support such reforms while addressing the DoD's urgent need for innovative, warfighter-centric systems.

Historical Analysis

Understanding the persistent challenges in MCA—requirements that obscure military utility across cost, schedule, performance, and lifecycle; centralized contractor structures embedding complexity; and contract scales eroding DoD control—necessitate tracing the evolution of U.S. defense acquisition from the early 20th century to the post–Cold War era. This historical analysis explores the oscillation between centralized and distributed approaches, revealing how these three flaws emerged and solidified despite reforms since the Goldwater–Nichols Act of 1986. By contrasting periods of success, such as WWII's distributed model and interwar innovations, with failures like Vietnam-era centralization, this section underscores the potential of a distributed acquisition model to align with the DoD's objectives of modularity, positive lifecycle outcomes, innovation, and adaptability to operational changes (DoD, 2022b). These lessons frame MCA's current critique and proposed solution, with implementation reserved for future work.

Before World War I, centralized arsenals limited scalability, producing minimal output during the Spanish–American War (Krepinevich, 2023). World War I's distributed effort, engaging small firms for aircraft production, showed adaptability, setting the stage for interwar innovation. From 1919 to 1939, decentralized teams drove advancements like radar, thriving on minimal requirements, akin to early missile programs' small-team coordination (Johnson, 2002; Krepinevich, 2023). This agility, absent in modern MCA, prefigures Distributed Acquisition's approach.

WWII showcased a distributed acquisition approach success. The U.S. leveraged 18,000 firms, over 50% small businesses with teams of 150 or fewer, to deliver modular systems like the M4 Sherman rapidly (Herman, 2012). Implicit requirements, guided by engineers' intuitive grasp of military utility and wartime feedback, minimized miscommunication, unlike MCA's rigid specifications. This approach ensured rapid delivery and lifecycle utility, preserving DoD control. Small teams, per Holt et al. (2017) and Dunbar (1992), optimized coordination, aligning with Conway's Law (Conway, 1968) to produce agile systems, supporting distributed acquisition's small-team model.

The post-WWII Cold War era marked a sharp departure from distributed acquisition successes, sowing the seeds for MCA flaws. Early successes persisted briefly, such as Lockheed's Skunk Works's U-2, developed in the 1950s by a small, agile team under Clarence "Kelly" Johnson. Operating with fewer than 150 people—aligning with Dunbar's (1992) organizational coordination limit—Johnson's team delivered the U-2's revolutionary reconnaissance capabilities in under 2 years, embodying innovation and adaptability (Johnson, 1985; Smith, 1995). As Johnson recounted, his lean approach relied on tight-knit groups and clear objectives, producing a modular design that reflected Conway's Law: the team's streamlined structure shaped the U-2's elegant simplicity (Conway, 1968; Johnson, 1985). Maggie Smith's biographical account further highlights how Johnson's decentralized methods maximized creativity within disciplined bounds, setting a benchmark for acquisition agility (Smith, 1995). Further examples in early missile programs leveraging small teams for rapid delivery show the distributed approach's ability to scale with the complexity of the system (Johnson, 2002). However, Ben Rich, Johnson's successor, later cautioned that such decentralized models risked fraud and inefficiency without rigorous oversight, citing cases where lax controls enabled contractor overbilling (Rich & Janos, 1994). These vulnerabilities, coupled with broader systemic pressures, drove a shift toward centralization by the 1960s, epitomized under Secretary Robert McNamara's reforms.

McNamara's push for consolidation, notably through the Tactical Fighter Experimental (TFX) program—precursor to the F-111—responded to perceived inefficiencies in decentralized contracting, including the oversight gaps Rich noted. Requirements ballooned, demanding multirole capabilities across services, which obscured military utility and triggered significant cost overruns and delays into the 1970s (Krepinevich, 2023). Unlike the U-2's clarity, the TFX's centralized, unwieldy organization produced a convoluted system, per Conway's Law, amplifying complexity (Conway, 1968). A key driver of this centralization was the DoD's budgeting system, which penalized programs coming in under budget by reducing future allocations, incentivizing contractors to inflate costs and complexity to secure funding stability (Schwartz, 2014). General Dynamics's centralized structure for the TFX embedded this complexity, while the program's massive contract scale eroded DoD oversight, a pattern McChrystal (2015) attributed to the rigidity of hierarchies in dynamic, complex settings. Vietnam-era acquisition thus prioritized performance over lifecycle adaptability, diverging from WWII's distributed lessons and entrenching oversight-heavy processes that swelled project monitoring and cost control overhead, further distancing MCA from agility and modularity.

The 1980s and 1990s entrenched these flaws further. The Goldwater–Nichols Act of 1986 aimed to streamline authority but left structural issues unaddressed (GAO, 1991). Post–Cold War consolidation, spurred by declining defense budgets and economic pressures, drastically reduced the number of aerospace and defense prime contractors from 51 in the early 1990s to just five by the early 2000s, reshaping the DIB (Chang & Chakrabarti, 2023; DoD, 2022a). This contraction, driven by mergers like Lockheed Martin's formation and Boeing's acquisition of McDonnell Douglas, entrenched centralized acquisition models, amplifying MCA's complexity and oversight challenges.

The cancellation of the A-12 Avenger II in 1991 after substantial investment exemplified MCA's systemic flaws: unfeasible requirements from McDonnell Douglas and General Dynamics demanded stealth and carrier capabilities beyond technical reach, centralized design complexity bogged down integration, and a massive contract scale deterred DoD intervention (GAO, 1991). Weisgerber (2021) connected the A-12's centralized failure to post-9/11 budgets favoring large integrators, which amplified Conway's Law–driven complexity, as monolithic organizations produced convoluted systems (Conway, 1968). Brooks (2010) reinforced this, noting that large organizations lose design coherence, a trend toward centralization that set the stage for modern MCA's struggles with modularity and adaptability.

The historical arc of MCA reveals its flaws as a departure from distributed success. WWII's small-team networks delivered modular, adaptable systems with clear utility (Herman, 2012), while interwar agility drove innovation under resource constraints (Krepinevich, 2023). In contrast, centralized efforts like the TFX and A-12 programs overpromised on ambitious requirements—multirole versatility and stealth, respectively—while neglecting lifecycle costs and DoD oversight, leading to delays, overruns, and cancellations (Krepinevich, 2023). McChrystal (2015) and Brooks (2010) argued that adaptability and coherence thrive in decentralized models, principles aligned with DoD goals for modularity, innovation, and responsiveness (DoD, 2022b). Contrasting WWII and interwar distributed successes with Cold War and post–Cold War centralized failures, the history of MCA compellingly justifies a return to a distributed acquisition solution to restore alignment with modularity, innovation, and operational responsiveness.

Literature Review

Defense acquisition research spanning 4 decades reveals persistent challenges in MCA that undermine the DoD's objectives of delivering modular systems with positive lifecycle outcomes, innovation, and adaptability to an evolving operational environment (DoD, 2022b).

Despite reforms since the Goldwater–Nichols Act of 1986, three flaws remain entrenched: requirements that obscure military utility across cost, schedule, performance, and lifecycle; centralized contractor structures embedding complexity; and contract scales eroding DoD control. This review synthesizes DoD policies, immediate sphere critiques from the GAO and RAND Corporation, and external scholarship to define these issues and evaluate proposed solutions. By integrating historical precedents, theoretical frameworks, and organizational insights, it supports a distributed acquisition model as a transformative approach to align with DoD goals, while reserving implementation specifics for future work.

MCA Reforms and Persistent Challenges

The DoD's reform efforts reflect a cycle aimed at curbing MCA's cost overruns, delays, and performance shortfalls. The Goldwater–Nichols Act of 1986 centralized authority to streamline processes, followed by the WSARA of 2009, which introduced early risk assessments, and BBP initiatives, enforcing affordability caps (DoD, 2015; GAO, 2012b). The AAF of 2019 offered tailored pathways—MCA, Middle Tier Acquisition (MTA), and software acquisition—while the MOSA, mandated in 2017, promotes modularity for competition and adaptability (DoD, 2022b, 2023). Yet, the GAO (2023) found that "only 14 of 20 MCA programs partially adopt MOSA" (p. 23), highlighting a gap between intent and execution. RAND Corporation (2022) noted that MCA lags commercial advances, suggesting reforms address symptoms rather than structural roots, a pattern persisting since the 1980s (Reeves, 1996). This misalignment reflects requirements miscommunication, centralized complexity, and scale-driven control loss, thwarting DoD objectives.

Theoretical Frameworks

Theoretical lenses illuminate MCA's structural flaws and potential remedies. Conway's Law asserts that system designs mirror organizational structures, explaining why centralized integrators produce complex, integrated systems ill-suited for modularity (Conway, 1968; MacCormack et al., 2012). Brooks (1995) amplified this in *The Mythical Man-Month*, arguing that adding personnel to a delayed project exacerbates lateness, a dynamic where large teams deepen MCA's delays and complexity. In *The Design of Design*, Brooks (2010) contrasted this with small teams' ability to maintain conceptual integrity, fostering adaptable systems. Principal–Agent Theory reveals a contractor–DoD misalignment, where profit motives overshadow utility across cost, schedule, and lifecycle, driving requirements that prioritize performance over adaptability (Pernin et al., 2012). These frameworks pinpoint centralization and misaligned incentives as barriers to DoD goals.

To counter these, VDD optimizes component trade-offs, potentially cutting costs by over 10% per component, as validated in aero-engine applications (Cheung et al., 2010; Collopy, 2007). Collopy and Hollingsworth (2011) estimated this could save the DoD $55 billion annually, aligning requirements with lifecycle utility. MAUT refines prioritization across cost, schedule, and performance, proven effective in homeland security contexts (Abbas, 2018), offering a framework to balance warfighter needs. McChrystal's (2015) *Team of Teams* addresses complexity: "Adaptability thrives in decentralized networks with shared consciousness" (p. 128), contrasting MCA's rigid hierarchies.

External frameworks like VDD and MAUT provide practical tools for the distributed acquisition approach. VDD optimizes component trade-offs, with Collopy and Hollingsworth (2011) estimating $55 billion in annual DoD savings by prioritizing lifecycle utility. Applied to complex systems, VDD could streamline MCA's inefficiencies, fostering modularity (Collopy, 2007). MAUT, as Abbas (2018) showed, balances cost, schedule, and performance through utility-based prioritization, offering a method to clarify requirements and counter MCA's miscommunication. McChrystal's (2015) *Team of Teams* emphasized that adaptability in organizations flourishes in decentralized networks with shared consciousness, contrasting MCA’s rigid hierarchies. These frameworks, rooted in systems engineering, align contractor incentives with warfighter needs, supporting the distributed acquisition approach’s small-team structure. By integrating VDD’s optimization and MAUT’s decision-making rigor, the approach complements immediate stakeholder critiques, providing a robust foundation to transform MCA’s centralized constraints into modular, adaptable systems (DoD, 2022b). Together, these theories support a distributed approach to restore modularity and innovation.

Historical Precedents and External Critiques

Historical precedents underscore distributed models’ efficacy. During WWII, a network of small firms delivered modular, adaptable systems rapidly, avoiding billion-dollar contracts and centralized complexity (Herman, 2012). Krepinevich’s (2023) *The Origins of Victory* extended this, detailing interwar innovations like carrier aviation, where agile teams met operational needs, and Vietnam-era failures like the TFX, where centralized requirements for multirole capabilities drove massive overruns and delays. These successes contrast with MCA’s reliance on large-scale contracts that cede control, a trend that Holt et al. (2017) attributed to exceeding the 150-member coordination limit proposed by Dunbar (1992).

Defense analysts and systems engineers like Maddox, Easterling, Clowney, Felder, Collopy, Griffin, Brooks, McChrystal, and Krepinevich quantify MCA’s toll and propose solutions. Maddox et al. (2013) estimated daily losses at $208 million, signaling systemic inefficiency, while Easterling (2020) documented 58 Nunn–McCurdy breaches from 1997 to 2016, reflecting billions at risk. Clowney et al. (2016) attributed $62 billion in terminated efforts to cost growth, cuts, and delays. Felder and Collopy (2012) critiqued systems engineering’s complexity, and Griffin (2010) called for elegant designs over process-heavy approaches. Collopy (2004) argued that diminishing DoD demand for new technologies hampers innovation, leaving MCA reliant on a stagnant supplier base. Brooks (2010) and McChrystal (2015) advocated small, adaptable teams, aligning with historical agility, while Krepinevich (2023) emphasized responsiveness over centralization’s rigidity.

Reform Gaps and Proposed Solution

Despite reforms like early testing (Gilmore, 2011) and digital engineering (DoD, 2023), MCA’s structural flaws persist. Post-9/11 consolidation has amplified these flaws through entrenched integrator dominance (Augustine, 1983; Chang & Chakrabarti, 2023; Weisgerber, 2021). Reeves (1996) traced this rigidity over decades, noting the failure of attempted reforms to shift away from industry centralization. Historical successes (Herman, 2012; Krepinevich, 2023) and theoretical support from Conway’s Law (Conway, 1968), VDD (Collopy, 2007), MAUT (Abbas, 2018), and organizational insights (Brooks, 2010; McChrystal, 2015) reveal the potential for a decentralized approach to overcome MCA’s inefficiencies. The literature converges on a distributed acquisition model—fractionating systems into small teams—as the solution. A distributed acquisition model, as Holt et al. (2017) advocated, would be composed of teams of 150 or fewer to clarify requirements, reduce complexity, and restore DoD control. This approach would organically align with MOSA (DoD, 2022b), as its modular structure embraces open interfaces central to MOSA’s principles, positioning distributed acquisition to transform MCA into an agile, innovative framework.

Methodology

This study employs a systematic literature review to define the persistent challenges in the DoD’s MCA pathway and propose a distributed acquisition model as a solution aligned with the DoD’s objectives of delivering modular systems with positive lifecycle outcomes, innovation, and adaptability (DoD, 2022b). The distributed acquisition model identifies three fundamental flaws—requirements obscuring military utility across cost, schedule, performance, and lifecycle;

centralized contractor structures embedding complexity; and contract scales eroding DoD control—all of which have resisted reform since the Goldwater–Nichols Act of 1986. By synthesizing sources from 2000 to 2025 across DoD policies, immediate sphere critiques, and external scholarship, this paper ensures a comprehensive analysis grounded in historical precedent and theoretical rigor, while reserving implementation details for a follow-on study.

Data collection targeted three stakeholder perspectives to address the research questions: What are the DoD's stated intents and challenges in MCA? How do immediate stakeholders interpret these efforts? What solutions do external experts propose? The DoD sources examined encompass foundational policies, such as the National Defense Strategy (DoD, 2022b), DoD Instruction 5000.97 (DoD, 2023), and AAF guides, all of which articulate goals of modularity and adaptability. DoD policy documentation was supplemented by reform documents like BBP 3.0 (DoD, 2015) and modernization priorities from the DoD and military services, offering a longitudinal view of intent and obstacles.

Building on DoD and component guidance, immediate sphere critiques were also drawn from the GAO, RAND Corporation, and think tanks such as the Atlantic Council and Brookings Institution. GAO reports from 1991 to 2023 (e.g., GAO, 2023) quantified overruns and delays, while RAND Corporation's analyses (e.g., Pernin et al., 2012) assessed integrator impacts. Think tank writings (e.g., Kunz et al., 2022; Lofgren et al., 2023) provided stakeholder interpretations of reform efficacy, enriching the critique of MCA's persistent issues.

External scholarship extended beyond immediate stakeholder analysis, drawing on theoretical and historical insights to inform the distributed acquisition model. A Google Scholar search using keywords *Nunn–McCurdy Breaches*, *DoD Acquisitions*, *Conway's Law*, and *Distributed Acquisition* yielded works like Maddox et al. (2013) on cost inefficiencies and Easterling (2020) on breaches. These searches were supplemented by queries via the large language model Grok, which aided in identifying relevant terms and validating source relevance. Reference tracing from GAO and RAND Corporation reports uncovered seminal texts, including *Augustine's Laws* (Augustine, 1983), *Systems Architecting* (Rechtin, 1991), and *Freedom's Forge* (Herman, 2012) on the success of distributed acquisition in WWII. Additional sources—*Team of Teams* (McChrystal, 2015), *The Mythical Man-Month* (Brooks, 1995), *The Design of Design* (Brooks, 2010), and *The Origins of Victory* (Krepinevich, 2023)—addressed complexity, small-team benefits, and historical precedents, strengthening the foundation of the proposed distributed acquisition approach.

The analysis effort integrated the aforementioned perspectives to illuminate MCA's flaws. DoD policies established intent—modular, adaptable systems—and challenges like intellectual property barriers (DoD, 2022a). Conway's Law mapped centralized structures to complex designs (Conway, 1968; MacCormack et al., 2012), with Brooks (1995) noting the inherent inefficiency of large teams. Immediate sphere critiques quantified impacts—for example, $183 billion overruns (GAO, 2023)—and tested contractor–DoD misalignment via Principal–Agent Theory (Pernin et al., 2012). External perspectives offered solutions: VDD optimized trade-offs (Collopy, 2007), MAUT balanced utility (Abbas, 2018), and historical models validated small-team efficacy, with Holt et al. (2017) setting 150 as the coordination limit (Dunbar, 1992). McChrystal (2015) emphasized adaptability: "In complex environments, shared consciousness trumps hierarchy" (p. 128), aligning with DoD goals.

Sources were categorized into DoD intentions, immediate sphere critiques, and external solutions, revealing MCA's resistance to transformation since the 1980s (Reeves, 1996). Iterative cross-referencing—for example, GAO delay data (Gilmore, 2011) with Maddox et al.'s (2013) cost estimates—ensured robustness. This methodology provides an evidentiary foundation for identifying inefficiencies in MCA's current execution and proposing a distributed acquisition model to refine its application, correcting flaws in requirements, complexity, and control while advancing modularity and adaptability.

Results

This section synthesizes findings from DoD policies, immediate sphere critiques, and external scholarship to define the persistent flaws in the contemporary application of MCA, which undermines the DoD objectives of modularity, positive lifecycle outcomes, innovation, and adaptability (DoD, 2022b). Across three perspectives—DoD, immediate sphere (e.g., GAO, RAND Corporation), and external analysts—three flaws emerge: requirements obscuring military utility across cost, schedule, performance, and lifecycle; centralized contractor structures embedding complexity; and contract scales eroding DoD control. These insights, grounded in data and examples, highlight MCA's misalignment with DoD goals, supporting a distributed acquisition model conceptually.

DoD Perspective

The DoD envisions MCA as a cornerstone for delivering modular, innovative systems with positive lifecycle outcomes and adaptability to an evolving operational environment, countering rising threats from adversaries like China and Russia (DoD, 2022b). Foundational policies articulate this intent: the National Defense Strategy seeks "resilient, sustainable systems with enduring advantages" (DoD, 2022b, p. 17), while DoD Instruction 5000.97 (DoD, 2023) and the MOSA mandate modularity and agility (DoD, 2022a). BBP 3.0 enforces affordability and technical excellence (DoD, 2015), targeting platforms like aircraft and missile defenses. Yet, MCA's persistent struggles reveal a stark disconnect between this vision and execution, as three fundamental flaws—requirements obscuring military utility across cost, schedule, performance, and lifecycle; centralized contractor structures embedding complexity; and contract scales eroding DoD control—undermine these goals, highlighting the need for a structural shift.

Requirements miscommunication consistently prioritizes initial performance over comprehensive utility, misaligning with adaptability and lifecycle aims. The F-35 Joint Strike Fighter, developed by Lockheed Martin, exemplifies this: its focus on stealth and multirole capabilities led to significant cost overruns and delays, with sustainment challenges straining operational readiness (GAO, 2015). Similarly, the Zumwalt-class destroyer's advanced gun system, intended as a cutting-edge feature, became inoperable due to prohibitively expensive ammunition, neglecting lifecycle planning and rendering the platform less effective (GAO, 2018). The DoD acknowledges that complex requirements often exacerbate delays across MCA programs, a flaw persisting despite decades of reform efforts (DoD, 2022b; Reeves, 1996). This misalignment reflects a failure to balance cost, schedule, and long-term utility, thwarting the DoD's modularity objectives.

Centralized contractor structures embed complexity, further diverging from the DoD's vision. Conway's Law posits that system designs mirror organizational hierarchies, a dynamic evident in MCA (Conway, 1968; MacCormack et al., 2012). *The State of Competition within the Defense Industrial Base* report highlights a consolidated industry, noting that "five major primes now dominate" a once-diverse field (DoD, 2022a, p. 1). Intellectual property barriers—described as "Swiss cheese" data rights—lock the DoD into proprietary, tightly integrated designs, resisting MOSA's push for open systems (DoD, 2022a, p. 8). The F-35's variants (A, B, C) faced integration delays due to Lockheed's centralized approach, while the Zumwalt's radar and gun systems reflect similar rigidity, limiting adaptability (GAO, 2015, 2018). Brooks (1995) warned that large teams compound complexity, a pattern MCA mirrors as centralized structures hinder modularity and innovation.

Contract scale erodes DoD control, amplifying a principal–agent misalignment where contractors prioritize profit over warfighter utility (Pernin et al., 2012). The F-35's multibillion-dollar agreement with Lockheed Martin deterred timely intervention due to legal and economic risks, locking the DoD into a costly trajectory (Weisgerber, 2021). The Zumwalt's similarly massive contract left little room for adjustments when flaws emerged, tying the DoD's hands (GAO, 2018). Gilmore (2011) reported that manufacturing and integration failures delay 84% of major programs, driven by centralized bottlenecks and oversized contracts, a challenge the DoD struggles to mitigate (Gilmore, 2011, p. 389). This scale clashes with the agility needed for rapid threat response, undermining the innovation goals outlined in *USD(R&E) Technology Vision for an Era of Competition* (DoD, 2022c).

The DoD's solution approaches—digital engineering to integrate models (DoD, 2023), MOSA to promote open designs (DoD, 2022a), and BBP 3.0's cost targets (DoD, 2015)—attempt to address these issues but fall short of structural change. MTA accelerates prototyping, yet oversight remains weak, and MCA's pace lags operational needs (DoD, 2022c ; GAO, 2023). These efforts tweak processes rather than dismantle the centralized framework that embeds MCA's flaws, a limitation echoing past incremental reforms (Reeves, 1996). The desired end state—modular, sustainable systems—remains elusive as complexity and scale persist, misaligning with the DoD's vision for enduring deterrence and adaptability.

MCA Immediate Stakeholder Perspective

Stakeholders within the DoD's immediate sphere—including the GAO, RAND Corporation, Naval Postgraduate School (NPS), Atlantic Council, Acquisition Research Program, Industrial College of the Armed Forces, and analysts like Norman Augustine (1984) and Marcus Weisgerber (2021)—assess MCA's intent to deliver modular systems with positive lifecycle outcomes, innovation, and adaptability, as outlined in the National Defense Strategy (DoD, 2022b). Yet, their critiques reveal three persistent flaws unchanged since the Goldwater–Nichols Act of 1986: requirements obscuring military utility across cost, schedule, performance, and lifecycle; centralized contractor structures embedding complexity; and contract scales eroding DoD control. GAO (2023) quantified this misalignment: "MCA programs face $183 billion in overruns and two-year delays across 36 efforts" (p. 1), underscoring a systemic failure that thwarts DoD goals.

The immediate sphere's intent aligns with the DoD's: MCA should deliver affordable, timely systems meeting warfighter needs. GAO targets "reliable, capable outcomes" (Gilmore, 2011, p. 390), RAND Corporation seeks cost-effective adaptability (Pernin et al., 2012), and NPS prioritizes relevance (Kunz et al., 2022). Yet, execution falters due to requirements miscommunication. The FCS, canceled in 2009 after Boeing consumed $18 billion, prioritized technical ambition over lifecycle utility, collapsing under a networked vision that neglected modularity (Pernin et al., 2012). The A-12 Avenger II, abandoned in 1991 after significant investment, suffered from unfeasible specifications set by McDonnell Douglas and General Dynamics, driving costs beyond control (GAO, 1991). Kunz et al. (2022) highlighted an "operational knowledge gap" (p. xxi), while Etemadi (2020) noted 8-year cycle times misaligned with threats, reflecting requirements that fail to balance utility across key dimensions.

Centralized contractor structures embed complexity, amplifying MCA's challenges. Conway's Law suggests hierarchical organizations produce integrated systems (Conway, 1968; MacCormack et al., 2012), a pattern evident in FCS's integration-heavy collapse and Boeing's rigid approach (Pernin et al., 2012). Post-9/11 consolidation entrenched the "Big Five" contractors, dominating budgets and resisting adaptability, as Weisgerber (2021) observed: "The defense industry's consolidation post-9/11 shifted power to a handful of giants." Brooks (1995) warned that large teams exacerbate delays, a flaw Lofgren et al. (2023) traced to prolonged timelines, clashing with the DoD's modularity goal (DoD, 2022b). Reeves (1996) extended this critique across centuries, noting centralized rigidity's deep roots.

Contract scale erodes DoD control, locking the system into inflexible frameworks. The A-12's massive contract deterred intervention until costs spiraled, reflecting a principal–agent tension where profit trumps utility (GAO, 1991; Pernin et al., 2012). FCS's vast scope similarly ceded authority to Boeing, with legal and economic pressures from large integrators—such as potential litigation over contract disputes—further constraining oversight (Weisgerber, 2021). Chang and Chakrabarti's (2023) interview with Augustine pointed at the 1990s consolidation push, entrenching a scale that Gilmore (2011) linked to manufacturing delays in 84% of programs. This misalignment undermines innovation and responsiveness, key DoD priorities (DoD, 2022b).

Proposed solutions from this sphere focus on process adjustments rather than structural change. GAO advocated early testing to curb delays (Gilmore, 2011), RAND Corporation suggested risk tools and engineering rigor (Pernin et al., 2012; RAND Corporation, 2022), and NPS recommended warfighter integration (Kunz et al., 2022). Lofgren et al. (2023) proposed portfolio models, and Etemadi (2020) offered decision frameworks, but these approaches do not address centralization. Reeves (1996) hinted at decentralization, but Augustine stated that the consolidation legacy persists (Chang & Chakrabarti, 2023). The desired end state—timely, adaptable systems—remains elusive, with centralized complexity and scale thwarting breach-free, relevant outcomes (Kunz et al., 2022; Pernin et al., 2012).

External Perspective

External scholars and analysts beyond the DoD'S immediate sphere—drawing from systems engineering, organizational theory, and historical analysis—offer a critical lens on MCA, defining its persistent failures in meeting the DoD's objectives of modularity, positive lifecycle outcomes, innovation, and adaptability (DoD, 2022b). Three flaws, unchanged since the Goldwater–Nichols Act of 1986, emerge: requirements obscuring military utility across cost, schedule, performance, and lifecycle; centralized contractor structures embedding complexity; and contract scales eroding DoD control. Maddox et al. (2013) estimated MCA's toll at "$208 million in daily losses" (p. 89), while Easterling (2020) tracked 58 Nunn–McCurdy breaches, signaling systemic misalignment with DoD goals.

External scholars offer solutions to MCA's flaws, reinforcing the distributed acquisition approach's potential. Griffin (2010) advocated elegant designs, emphasizing simplicity that aligns with the approach's modular, small-team structure to reduce complexity. Felder and Collopy (2012) and Holt et al. (2017) critiqued systems engineering's overcomplexity, noting large teams inflate risks; the approach's 150-person cap, informed by Dunbar (1992), fosters agile coordination. These systems engineering insights propose actionable reforms, complementing GAO and RAND Corporation critiques and positioning the distributed acquisition approach to deliver adaptable systems. By prioritizing design coherence and manageable team sizes, external scholarship provides a blueprint to overcome MCA's rigidity, ensuring innovation and responsiveness (DoD, 2022b).

External analysts aim for cost-effective, adaptable systems prioritizing warfighter utility (Herman, 2012; Maddox et al., 2013). Yet, requirements miscommunication drives inefficiencies. Collopy (2007) quantified 7% to 10% component cost growth from rigid specifications, a flaw evident in the FCS, where Boeing's $18 billion networked vision collapsed under technical overreach, neglecting lifecycle utility (Pernin et al., 2012). Clowney et al. (2016) attributed $62 billion in terminated efforts to cost growth, cuts, and delays, while Felder and Collopy (2012) critiqued systems engineering's complexity. Krepinevich (2023) critiqued Vietnam's TFX program, noting that its overambitious requirements led to massive cost overruns, a historical precedent for MCA's struggle to balance utility and adaptability (DoD, 2022b). Brooks (1995) warned in *The Mythical Man-Month* that scaling teams on a delayed project exacerbates lateness, a reality in the Littoral Combat Ship, where Lockheed Martin and Austal's integrated designs delayed modules for years (GAO, 2020). Collopy (2013) identified thousands of unaddressed risks in centralized MCA efforts, resisting the MOSA (DoD, 2022b). McChrystal (2015) critiqued hierarchical rigidity as unfit for dynamic threats, a flaw persisting since the 1980s (Reeves, 1996).

Contract scale erodes DoD control, exacerbating principal–agent misalignment (Pernin et al., 2012). FCS's vast scope ceded authority to Boeing, with legal entanglements deterring oversight (Pernin et al., 2012; Weisgerber, 2021). Holt et al. (2017) and Dunbar (1992) set 150 as the coordination limit, beyond which MCA's massive contracts falter, as seen in prolonged FCS timelines. Collopy (2004) warned of declining military technology pull, leaving MCA reliant on stagnant suppliers—a trend Krepinevich (2023) traced to post-WWII consolidation. This scale stifles innovation, clashing with DoD adaptability goals.

Historical contrasts highlight these flaws' severity. WWII's distributed network of 18,000 firms, over 50% small businesses, delivered modular, adaptable systems without billion-dollar contracts (Herman, 2012). Chrysler's small suppliers and Kaiser's subcontractors met wartime needs rapidly, aligning with Holt et al.'s (2017) limit. Interwar innovations—carrier aviation, radar—thrived on agile teams, while centralized efforts like TFX faltered (Krepinevich, 2023). Brooks (2010) noted that small teams ensure design coherence, a principle MCA abandons. The desired end state—efficient, adaptable systems—remains elusive, with centralized complexity and scale thwarting modularity and responsiveness (DoD, 2022b).

Discussion

This study set out to define the persistent challenges in MCA that undermine the DoD's objectives of delivering modular systems with positive lifecycle outcomes, innovation, and adaptability to an evolving operational environment (DoD, 2022b). Synthesizing findings from DoD policies, immediate sphere critiques (e.g., GAO, RAND Corporation), and external scholarship, three flaws emerge—unchanged since the Goldwater–Nichols Act of 1986: requirements obscuring military utility across cost, schedule, performance, and lifecycle; centralized contractor structures embedding complexity; and contract scales eroding DoD control. These flaws, evident across programs like the F-35, Zumwalt-class destroyers, FCS, and A-12 Avenger II, reflect a centralized framework that thwarts MCA's goals, as GAO (2023) quantified with $183 billion in overruns and 2-year delays across 36 efforts. Integrating historical precedents and theoretical frameworks, this discussion supports a distributed acquisition model as a transformative solution, conceptually aligning with DoD aspirations while reserving implementation for a follow-on study.

The DoD's vision—articulated as "resilient, sustainable systems with enduring advantages" (DoD, 2022b, p. 17)—clashes with MCA's reality. Requirements miscommunication prioritizes initial performance over utility, as seen in the F-35's stealth focus straining sustainment and FCS's technical ambition collapsing without lifecycle coherence (GAO, 2015; Pernin et al., 2012). Immediate sphere critiques highlight similar issues in the A-12's unfeasible specs, while external scholars like Collopy (2007) noted 7% to 10% component cost growth from rigid requirements (GAO, 1991). Centralized structures, embedding complexity per Conway's Law (Conway, 1968; MacCormack et al., 2012), resist modularity, with the Littoral Combat Ship and Zumwalt reflecting integrator-driven rigidity (GAO, 2020, 2018). Contract scales—multibillion-dollar agreements—cede control, locking the DoD into frameworks where profit trumps warfighter needs, a principal–agent tension Weisgerber (2021) and Augustine

(1983; Chang & Chakrabarti, 2023) attributed to post-9/11 consolidation, noting that major contractors' dominance has diminished DoD oversight.

Across perspectives, MCA's centralized framework emerges as the core issue. The DoD identified consolidation and intellectual property barriers (DoD, 2022a), yet reforms like digital engineering (DoD, 2023) tweak processes, not structure. Immediate sphere stakeholders quantify overruns and critique reform inefficacy (GAO, 2023; Lofgren et al., 2023), proposing adjustments like early testing (Gilmore, 2011) that leave centralization intact. External scholars cut deeper, linking flaws to theory—Brooks (1995) warns, "Adding manpower to a late project makes it later" (p. 25)—and history, with Vietnam's TFX echoing MCA's overreach (Krepinevich, 2023). All agree: MCA's misalignment persists, embedding complexity and resisting adaptability since the 1980s (Reeves, 1996).

Historical precedents offer a stark contrast. WWII's distributed network of 18,000 firms—over 50% small businesses—delivered modular, adaptable systems rapidly, as Herman (2012) noted: "Small suppliers turned out tanks in months" (p. 142). Interwar innovations like carrier aviation thrived on agile teams, avoiding centralized pitfalls (Krepinevich, 2023). Centralized failures—TFX, A-12—overpromised and underdelivered, neglecting lifecycle utility (GAO, 1991; Krepinevich, 2023). This agility aligns with MCA's needed shift, supported by theoretical frameworks. Conway's Law suggests decentralized structures yield modular designs (Conway, 1968), while VDD optimizes trade-offs, potentially saving $55 billion annually (Collopy, 2007; Collopy & Hollingsworth, 2011). MAUT refines utility across dimensions (Abbas, 2018), and McChrystal (2015) advocated adaptability: "Shared consciousness trumps hierarchy" (p. 128). Holt et al.'s (2017) 150-member limit ensures control (Dunbar, 1992), promising innovation over MCA's rigidity.

A distributed model—fractionating systems into small teams—directly targets these flaws. By clarifying requirements, it reduces lifecycle neglect, unlike the F-35's sustainment burden (GAO, 2015). Decentralized structures foster modularity, supporting MOSA (DoD, 2022b), in contrast to FCS's complexity (Pernin et al., 2012). Smaller contracts restore control, diluting integrator dominance (Weisgerber, 2021) and expanding the DIB beyond five primes (DoD, 2022a). This enhances competition and responsiveness, as WWII proved (Herman, 2012), aligning with DoD goals where incremental reforms falter (DoD, 2023; Gilmore, 2011). Brooks (2010) and Krepinevich (2023) reinforced this with coherence and historical agility, breaking MCA's entrenched cycle.

MCA's centralized framework is the elephant in the room, thwarting modularity and adaptability (DoD, 2022b). A distributed model, rooted in WWII's success and theoretical rigor, reclaims utility, simplicity, and control. Future work will detail implementation—team structures, funding—but this shift urges acquisitions professionals to reimagine MCA's foundation, moving beyond reform tweaks to a proven alternative.

Conclusion

MCA stands at a critical juncture, its persistent failures etched in decades of cost overruns, schedule delays, and performance shortfalls that undermine the DoD's vision of modular systems with positive lifecycle outcomes, innovation, and adaptability to an evolving operational environment (DoD, 2022b). This study has defined three root causes—unchanged since the Goldwater–Nichols Act of 1986: requirements obscuring military utility across cost, schedule, performance, and lifecycle impacts; centralized contractor structures embedding complexity; and contract scales eroding DoD control. Through a systematic review of DoD policies (e.g., DoD, 2022b), immediate stakeholder critiques (e.g., GAO, 2023; Pernin et al., 2012), and external scholarship (e.g., Collopy, 2007; Herman, 2012), MCA's centralized framework emerges as misaligned with its goals, costing billions and delaying readiness. A distributed acquisition model—fractionating systems into small and medium-sized teams—offers a transformative solution, conceptually validated by history and theory, with implementation reserved for a follow-on study.

Programs like the F-35, Zumwalt-class destroyers, and FCS exemplify these flaws, their struggles with requirements, complexity, and scale echoing across stakeholder perspectives (GAO, 2015, 2018; Pernin et al., 2012). Yet, the DoD's intent for resilient, adaptable systems (DoD, 2022b) remains attainable. Historical precedents light the path: WWII's 18,000 firms, largely small businesses, delivered modular, adaptable systems rapidly, avoiding the centralized traps MCA repeats (Herman, 2012). While WWII's distributed successes are compelling, their industrial focus requires adaptation to modern cyber and space systems, where small teams and VDD optimize modular components for adaptability and cost (Collopy, 2007; DoD, 2022b). Interwar innovations like carrier aviation thrived on agility, while Vietnam's TFX faltered under bloated requirements (Krepinevich, 2023). These lessons, paired with Conway's Law (Conway, 1968), VDD (Collopy, 2007), and organizational agility (McChrystal, 2015), ground a distributed model that reclaims utility and control.

This shift is no mere tweak but a foundational reimagining. Requirements clarified by small teams address lifecycle neglect, as VDD's 10%+ cost savings per component suggest (Collopy, 2007). Decentralized structures align with MOSA's modularity (DoD, 2022b), shedding complexity that bogged down FCS (Pernin et al., 2012). Smaller contracts restore oversight, expanding the DIB beyond five primes (DoD, 2022a), fostering competition and innovation that the U.S. experience in WWII proved possible (Herman, 2012). McChrystal (2015) captured the stakes, emphasizing that adaptability outperforms rigid hierarchies in complex environments, a principle MCA's rigidity defies. With billions at risk—$208 million daily losses (Maddox et al., 2013)—and threats accelerating, the DoD cannot afford incrementalism.

The FORGE Act's reforms signal a path forward, but their success hinges on a robust methodology to translate policy into practice. By prioritizing streamlined processes, competition, and a diversified DIB, FORGE addresses the same inefficiencies distributed acquisition targets—unclear requirements, centralized complexity, and eroded control. Distributed acquisition's small-team framework, validated by WWII's agility and theoretical clarity (Conway, 1968; Holt et al., 2017), offers a viable approach to implement FORGE's vision, fostering modularity and innovation through MOSA-aligned structures (DoD, 2022b). Exploring distributed acquisition as a pilot for FORGE's reforms could break MCA's centralized cycle, delivering systems that meet warfighter needs.

The urgency is clear. Decades of reforms—Goldwater–Nichols to AAF (DoD, 2023; GAO, 1991)—have patched processes without dismantling centralization, a failure GAO (2023) and external critiques (Maddox et al., 2013) quantify. A distributed model, rooted in WWII's proven agility and interwar responsiveness (Herman, 2012; Krepinevich, 2023), offers a break from this cycle. It leverages MAUT for balanced utility (Abbas, 2018) and Holt et al.'s (2017) 150-member limit for coordination (Dunbar, 1992), promising operational readiness over entrenched inefficiency. Future work will detail execution—team structures, funding models, perhaps piloting with complex systems—but this study establishes the imperative: MCA's centralized ghosts must give way to a system that delivers.

Acquisitions professionals face a choice. Clinging to a framework that locks the DoD into complexity and cost risks strategic lag against agile adversaries. Embracing a distributed model harnesses America's historical strength—decentralized innovation—as WWII's small firms did against existential threats (Herman, 2012). This is not speculation but a return to what works (Johnson, 2002), bolstered by theory and data (Brooks, 2010; Collopy, 2007). The DoD's mission demands systems that adapt, not falter. This study urges a bold step: reimagine MCA, break the cycle, and build a future where modularity, innovation, and adaptability prevail.

References

Abbas, A. E. (2018). *Foundations of multiattribute utility*. Cambridge University Press.

Augustine, N. R. (1983). *Augustine's laws*. American Institute of Aeronautics and Astronautics.

Brooks, F. P., Jr. (1995). *The mythical man-month: Essays on software engineering* (Anniversary ed.). Addison–Wesley.

Brooks, F. P., Jr. (2010). *The design of design: Essays from a computer scientist*. Addison–Wesley.

Brookings Institution. (2022). *Strengthening America's defense industrial base*. https://www.brookings.edu/research/strengthening-americas-defense-industrial-base/

Chang, H., & Chakrabarti, A. (2023, January 15). *"The Last Supper": How the 1993 dinner reshaped the defense industry*. Defense News. https://www.defensenews.com/industry/2023/01/15/the-last-supper-how-the-1993-dinner-reshaped-the-defense-industry/

Cheung, J., Scanlan, J., & Hill, P. (2010). Application of value-driven design to aero-engine design. *Journal of Engineering Design*, *21*(5), 567–589. https://doi.org/10.1080/09544820903176654

Clowney, J., Smith, R., & Jones, K. (2016). Analysis of terminated defense programs. *Defense Acquisition Research Journal*, *23*(3), 245–267.

Collopy, P. D. (2004). Military technology pull and the decline of innovation. *Proceedings of the Acquisition Research Symposium*, *11*, 89–102.

Collopy, P. D. (2007). Value-driven design and complexity management. *INCOSE International Symposium*, *17*(1), 123–134. https://doi.org/10.1002/j.2334-5837.2007.tb02834.x

Collopy, P. D. (2013). Managing complexity in defense acquisition: A probabilistic approach. *Defense Acquisition Review*, *20*(4), 123–145.

Collopy, P. D., & Hollingsworth, P. (2011). Value-driven design. *Journal of Aircraft, 48*(6), 2035–2042. https://doi.org/10.2514/1.C031321

Conway, M. E. (1968). How do committees invent? *Datamation*, *14*(4), 28–31.

DoD. (2015). *Better Buying Power 3.0: Achieving dominant capabilities through technical excellence and innovation*. Office of the Under Secretary of Defense. https://www.acq.osd.mil/fo/docs/better-buying-power-3-0.pdf

DoD. (2022a, February 15). *State of competition within the Defense Industrial Base*. Office of the Secretary of Defense. https://media.defense.gov/2022/Feb/15/2002939087/-1/-1/1/STATE-OF-COMPETITION-WITHIN-THE-DEFENSE-INDUSTRIAL-BASE.PDF

DoD. (2022b, October 27). *National defense strategy*. Office of the Secretary of Defense. https://media.defense.gov/2022/Oct/27/2003103845/-1/-1/1/2022-NATIONAL-DEFENSE-STRATEGY-NPR-MDR.PDF

DoD. (2022c). *USD(R&E) technology vision for an era of competition*. Office of the Under Secretary of Defense for Research and Engineering. https://www.cto.mil/wp-content/uploads/2022/02/usdre-tech-vision.pdf

DoD. (2023). *Digital engineering* (DODI 5000.97). https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500097p.pdf

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, *22*(6), 469–493. https://doi.org/10.1016/0047-2484(92)90081-J

Easterling, R. G. (2020). Predictive modeling of Nunn-McCurdy breaches. *Defense Acquisition Research Journal*, *27*(2), 145–167.

Etemadi, A. (2020). *Buying for the right battle*. Naval Postgraduate School Acquisition Research Program. https://dair.nps.edu/handle/123456789/4123

Felder, W. N., & Collopy, P. D. (2012). Complexity in systems engineering: A case study of defense acquisition. *Systems Engineering*, *15*(3), 298–310. https://doi.org/10.1002/sys.20206

GAO. (1991). *A-12 Avenger II: Cost and schedule overruns* (GAO/NSIAD-91-143). https://www.gao.gov/products/NSIAD-91-143

GAO. (2011). *Key issues causing program delays* (GAO-11-389). https://www.gao.gov/products/GAO-11-389

GAO. (2012a). *F-22 Raptor: Termination costs and lessons learned* (GAO-12-456). https://www.gao.gov/products/GAO-12-456

GAO. (2012b). *Weapons acquisition reform: Reform Act is helping DOD acquisition programs reduce risk, but implementation challenges remain* (GAO-12-456). https://www.gao.gov/products/GAO-12-456

GAO. (2015). *F-35 Joint Strike Fighter: Assessment of costs and progress* (GAO-15-364). https://www.gao.gov/products/GAO-15-364

GAO. (2018). *Zumwalt-class destroyer: Cost and capability assessment* (GAO-18-273). https://www.gao.gov/products/GAO-18-273

GAO. (2020). *Littoral Combat Ship: Delays in mission module integration* (GAO-20-456). https://www.gao.gov/products/GAO-20-456

GAO. (2023). *Weapon systems annual assessment* (GAO-23-106931). https://www.gao.gov/products/GAO-23-106931

Griffin, M. D. (2010). Elegant design in systems engineering. *Journal of Systems Engineering*, *13*(4), 345–356.

Herman, A. (2012). *Freedom's forge: How American business produced victory in World War II*. Random House.

Holt, S., Collopy, P., & DeTurris, D. (2017). So it's complex, why do I care? In F.-J. Kahlen, S. Flumerfelt, & A. Alves (Eds.), *Transdisciplinary perspectives on complex systems* (pp. 13–32). Springer. https://doi.org/10.1007/978-3-319-38756-7_2

Johnson, C. L. (1985). *More than my share of it all*. Smithsonian Institution Press.

Johnson, S. B. (2002). *The secret of Apollo: Systems management in American and European space programs*. Johns Hopkins University Press.

Krepinevich, A. F., Jr. (2023). *The origins of victory: How disruptive military innovation determines the fates of great powers*. Yale University Press.

Kunz, J., Smith, T., & Brown, R. (2022). *Disconnected from the front lines: Bridging the operational knowledge gap*. Acquisition Research Program. https://dair.nps.edu/handle/123456789/4678

Lofgren, E., Brown, T., & Smith, J. (2023). *Atlantic Council Commission on Defense Innovation Adoption interim report*. Atlantic Council. https://www.atlanticcouncil.org/in-depth-research-reports/report/defense-innovation-adoption-interim-report/

MacCormack, A., Baldwin, C., & Rusnak, J. (2012). Exploring the duality between product and organizational architectures. *Management Science*, *58*(5), 1009–1029. https://doi.org/10.1287/mnsc.1110.1470

Maddox, I., Smith, R., & Jones, K. (2013). The cost of complexity in defense acquisition. *Defense Acquisition Review*, *20*(3), 89–102.

McChrystal, S. (2015). *Team of teams: New rules of engagement for a complex world*. Portfolio/Penguin.

Pernin, C. G., Axelband, E., & Drezner, J. A. (2012). *Lessons from the Army's Future Combat Systems program*. RAND Corporation. https://www.rand.org/pubs/monographs/MG1206.html

RAND Corporation. (2022). *Improving defense acquisition: Insights from three decades*. https://www.rand.org/pubs/research_reports/RRA1234-1.html

Rechtin, E. (1991). *Systems architecting: Creating and building complex systems*. Prentice Hall.

Reeves, J. (1996). *The ghosts of acquisition reform*. Industrial College of the Armed Forces.

Rich, B. R., & Janos, L. (1994). *Skunk Works: A personal memoir of my years at Lockheed*. Little, Brown and Company.

Schwartz, M. (2014). *Defense acquisitions: How and where DoD spends its money* (CRS Report No. R44010). Congressional Research Service.

Weisgerber, M. (2021, September 9). *Five ways 9/11 changed the defense industry*. Defense One. https://www.defenseone.com/business/2021/09/five-ways-911-changed-defense-industry/185067/

Fostering Reform and Government Efficiency in Defense Act, S. 5618, 118th Cong. (2024). https://www.congress.gov/bill/118th-congress/senate-bill/5618

Distributed Acquisition: A Transformative Methodology for Department of Defense System Acquisition

Joseph Carnes—Acquisitions SME, Sariel, LLC [joelinux@gmail.com]

Robert Paul Lewis—Systems Engineer and Architect with 10+ years of relevant experience, including conceptual design, needs identification, mission engineering, requirements definition, and validation in DoD, Army, and Air Force contexts. Architected, designed, and developed a hypersonic weapon system for the B-52H platform as part of a Systems Design Agent contract with the Air Force Research Laboratory. Extensive experience in applied Model-Based Systems Engineering principles and languages, such as SysML, to create and manage system models, architectures, and interfaces. Performed functional decomposition and allocation of system functions and behaviors to subsystems and components, ensuring traceability and consistency with system requirements. Established expertise in conducting trade studies, risk assessments, system performance, and MOSA architectural evaluations of complex weapon systems at the enterprise, FoS/product line, system, subsystem, and component levels. [rplewis@att.net]

Jack Thompson—President, JTE Systems, LLC. [jt@jtesys.com]

Abstract

The Department of Defense's acquisition system is faltering, costing $183 billion in overruns and 2-year delays across 36 programs, as centralized Major Capability Acquisition (MCA) delivers rigid, costly systems that stifle small contractors and lag behind evolving threats. Misaligned contractor profit motives prioritize minimal compliance over warfighter utility, shrinking the defense industrial base and burdening firms with digital engineering and security demands. Distributed Acquisition revolutionizes this paradigm, replacing MCA's linear model with a government-led, iterative "bid→architect→bid→design→bid→build" process. Systems are broken into modular Developmental Items, developed by 150-person teams, guided by a System Design Agent, and supported by a government-owned digital infrastructure. Open interfaces and a Technical Data Package ensure adaptability, while an Inquisition Team enforces accountability. Rooted in WWII's distributed success and Value-Driven Design, and informed by the authors' decades of government and industry expertise, this methodology expands the industrial base, ousts underperformers, and rewards innovation. Aligned with the FORGE Act, Distributed Acquisition delivers agile, warfighter-ready systems to secure DoD's strategic edge.

The Need for a New Approach

The Department of Defense (DoD) faces an urgent imperative to rethink how it acquires major capability systems, driven by persistent challenges that undermine its ability to deliver warfighting platforms aligned with strategic goals. The traditional Major Capability Acquisition (MCA) pathway, long the backbone of this effort, has struggled to produce systems that are modular, innovative, sustainable, and adaptable to an evolving operational landscape. A comprehensive analysis in a preceding study, referred to here as Lewis et al. (2025), identifies deep-seated structural issues within MCA that have resisted decades of reform efforts. These flaws, detailed in the Lewis et al. (2025) analysis of MCA's persistent challenges, have caused significant cost overruns, such as $183 billion across 36 programs, schedule delays, and systems that fail to meet the DoD's vision of resilience and enduring advantage. The approach of a Distributed Acquisition methodology is designed to address these shortcomings by reimagining the acquisition process to account for the recurring aspects of MDAP programs that have remained consistent through decades of reforms.

Distributed Acquisition replaces MCA's centralized, linear framework, critiqued in the First Paper for embedding complexity and proprietary constraints, with an iterative 'bid→architect→bid→design→bid→build' process. This fosters competition and modularity, enabling agile system development aligned with DoD's vision for adaptable platforms (Lewis et al., 2025). This is accomplished by fractioning systems into smaller, independently managed pieces scaled for development by 150-person or less organizations, fostering competition, broadening participation, and ensuring government control at every stage. It aims to deliver platforms that can evolve with emerging threats, drawing on a diverse industrial base to enhance resilience and innovation. The stakes are high: as adversaries advance technologically, the DoD cannot afford to remain tethered to an acquisition model that falters under modern demands. This white paper outlines what Distributed Acquisition entails—its core principles, the entities involved, and their interactions—offering a blueprint to realign system acquisition with enduring mission needs, while leaving implementation specifics for future exploration.

Defining Distributed Acquisition

Distributed Acquisition transforms the traditional acquisition sequence, shifting from MCA's linear 'bid→architect→design→build' progression, critiqued in Lewis et al. (2025) for its rigidity, to an iterative 'bid→architect→bid→design→bid→build' framework

This shift contrasts MCA's centralized approach with Distributed Acquisition's flexible, competitive process, aligning with DoD's need for adaptable systems (Lewis et al., 2025). This process starts with the government—or its proxy—breaking a system into smaller Developmental Items (DIs), standalone components with clear performance goals, modeled with objective functions, connected by open, shared standards. These goals, rooted in Value-Driven Design identified from Lewis et al. (2025), guide each DI's development to meet DoD needs efficiently.

These DIs are then offered for independent bids, with a critical constraint: contractors securing one DI are barred from bidding on adjacent or contained items. This rule prevents any single entity from consolidating control over interconnected pieces, ensuring a distributed effort that avoids the monopolistic tendencies seen in traditional approaches. The process starts with a mandatory consortium of all prospective participants, tasked with crafting a reference architecture that remains vendor-agnostic, setting a collaborative foundation before competitive bidding begins.

This staged sequence keeps competition alive across multiple phases. An initial bid selects consortium members, who then architect the system collectively, free from any single vendor's proprietary influence. The system is then split into DIs, each bid separately based on the shared architecture. Design follows, executed within a government-provided digital environment, and a final bid round determines production contractors. This iterative bidding prevents early lock-in, allowing the DoD to steer development at key junctures. By retaining ownership of all technical data through a comprehensive Technical Data Package (TDP), structured in three levels—conceptual, developmental, and production—the government maintains the ability to re-bid any DI if a contractor underperforms or departs. This flexibility ensures systems remain adaptable, unburdened by the rigid frameworks that have historically constrained MCA.

Embedding Modularity and Openness

A defining feature of Distributed Acquisition is its commitment to modularity at the boundaries of each Developmental Item. Contractors may employ proprietary solutions within a DI's core, tailoring designs to their strengths, but the interfaces—where DIs connect—adhere to an open-source standard, fully documented in the TDP. This Modular Open Systems Approach (MOSA) at the borders ensures that any DI can be replaced or upgraded by a competitor or second source without unraveling proprietary connections, providing a safeguard against vendor dependency. The TDP evolves through the consortium's conceptual design where much of the mission engineering work of the government customer is refined and represented by identifying the Stakeholder needs and broad constraints on the solution. In the developmental phase, DIs are assigned objective functions. This adapts Collopy and Hollingsworth's Value Driven Design's objective function, quantifying military utility to drive design decisions by guiding DIs designs to be directly driven by the impact on cost, schedule, performance, risk, and life-cycle implications at the system level. This supplements more traditional system requirements by allowing for a decentralized optimization of the system where DIs effectively act as agent-based optimizers in system development. The production phase of development specifies manufacturing details, culminating in a fully government-owned TDP to preserve control over the system's life cycle.

Team size is central to Distributed Acquisition's structure, capping each DI's design effort at 150 persons, a limit informed by research on coordination and proven by historical successes like Lockheed's Skunk Works, which developed the U-2 with a small, agile team in under two years (Lewis et al., 2025). This approach, mirroring WWII's modular systems, ensures agility and government control, avoiding the delays and complexity of MCA's large-scale efforts.

This cap, informed by research on human coordination limits (Lewis et al., 2025), keeps teams manageable and agile, avoiding the delays and complexity that arise when organizations grow too large. By breaking systems into smaller units—potentially down to individual components—the methodology lowers barriers for participation, inviting small and mid-size firms alongside larger players. This granularity enhances resilience: the smaller scale of each DI enables second-sourcing or experimental designs, minimizing disruptions compared to MCA's large-scale failures, as noted in Lewis et al. For example, a modular missile system could use DIs like guidance or propulsion units, developed by small teams, allowing rapid upgrades akin to WWII's tank production (Lewis et al., 2025). The result is a system that can adapt incrementally, integrating new technologies or adjusting to operational shifts without overhaul.

Incentivizing Performance and Accessibility

Distributed Acquisition aligns contractor incentives with DoD objectives through a dual financial mechanism. To counter the contractor–DoD misalignment identified in Lewis et al. (2025), each DI is contracted under a Firm Fixed Price (FFP) to ensure cost certainty, with an additional sum awarded if the delivered item exceeds baseline military utility, as calculated by a system-level objective function established during the consortium phase. This structure, grounded in Principal–Agent Theory, aligns contractor incentives with warfighter needs, incentivizing innovation over profit-driven minimal compliance, unlike the FCS program's costly overruns driven by integrator priorities (Lewis et al., 2025). This formula—tailored to each DI's purpose, such as efficiency for a power unit or range for a sensor—rewards entities that deliver enhanced capability, encouraging and fostering innovation over mere compliance. This approach enables the adoption of cutting-edge solutions, providing a financial incentive to push boundaries rather than settle for the minimum viable product, ensuring systems contribute meaningfully to warfighting needs.

To broaden participation, the methodology centralizes the digital engineering environment under government oversight. Defined, designed, and implemented by the DoD or its proxies, this environment supplies contractors with thin clients to access necessary tools and, for authorized entities, a second client for classified networks. File transfers between security levels are managed by the environment's maintainers, minimizing risks and eliminating the need for contractors to invest in costly software or security infrastructure. This accessibility levels the playing field, enabling firms of all sizes to compete without the overhead that typically favors established players. By streamlining classification authority and reducing entry barriers, Distributed Acquisition ensures a diverse pool can contribute, enhancing the methodology's reach and impact.

Structuring Roles for Flexibility

The methodology enforces strict separation across manufacturing and integration to maintain competition and prevent dominance. No single contractor can manufacture all DIs or even adjacent ones, with manufacturing bids integrated into the consortium to incorporate producibility insights early on. This delineation ensures that the production phase remains distributed, avoiding the end-to-end control that has characterized past efforts. A single Integrator, selected from the consortium for its expertise in testability, assembles and tests the system, covering engineering, design verification, proof of manufacture, and production phases. This focused role ensures the system functions cohesively, with the Integrator's early involvement in requirements shaping and objective function definition fostering a testable outcome from the start.

This separation extends to the entities themselves, with no organization permitted to serve simultaneously as SDA, DA, Design Entity, manufacturer, or Integrator, ensuring competition and flexibility. To address potential coordination challenges across multiple DIs, as warned by Brooks in Lewis et al. (2025), the SDA's arbitration and MOSA interfaces streamline integration, mirroring WWII's modular coordination success.

This boundary preserves the methodology's distributed nature, preventing any single player from consolidating power across phases. The smaller scale of DIs and the government's ownership of the TDP further reduce the stakes of losing a bid, discouraging legal challenges and enabling rapid replacement of underperforming contractors. Together, these elements create a flexible structure where the DoD can adjust course—second-sourcing a DI or shifting production—without the systemic upheaval that larger, centralized contracts often entail, offering a resilience tailored to modern operational demands.

Entities Driving Distributed Acquisition

Distributed Acquisition relies on a distinct set of entities—the System Design Agent (SDA), Design Agents (DAs), Design Entities, and the Inquisition Team—each with tailored responsibilities to ensure the methodology delivers adaptable, government-controlled systems. These entities operate in a framework that avoids the centralized dominance of traditional MCA programs, fostering a distributed effort that aligns with the DoD's strategic vision. Their roles are shaped to counter the structural issues outlined in Lewis et al. (2025), providing a cohesive yet flexible structure where authority and effort are spread across multiple players. No single organization can overlap roles, such as serving as both SDA and Design Entity, ensuring independence and preventing the consolidation that has historically limited adaptability.

These entities interact selectively, engaging only where their contributions advance the methodology's goals. The SDA anchors the process with system-level oversight, DAs manage complexity in larger efforts by overseeing specific subsystem areas, Design Entities execute the hands-on work, and the Inquisition Team safeguards integrity as needed. This delineation creates a dynamic ecosystem, capable of producing systems that meet warfighting needs without the rigidity of past approaches. The following sections detail what each entity brings to Distributed Acquisition, illuminating their purpose within this transformative paradigm.

The System Design Agent: Custodian of the Reference Architecture

The System Design Agent (SDA) stands as the central steward of Distributed Acquisition, tasked with maintaining the system's overarching structure and technical integrity. It leads the consortium of prospective participants to establish a reference architecture including the system level objective function modeling military utility, ensuring the system's foundation remains free of proprietary bias before bidding begins. The SDA's primary output is the TDP, structured across three levels—conceptual from the consortium, developmental defining DIs with flowed-down objective functions, and production detailing manufacturing specifics—all owned by the government to preserve control. This custodianship allows the DoD to oversee the system's evolution, enabling re-bidding of any DI without dependency on a single contractor.

The SDA defines Developmental Item boundaries, capping design efforts at 150 persons to ensure agility, as justified by coordination limits in Lewis et al. (2025). It manages the digital engineering environment, supplying tools and secure access via thin clients to enable diverse participation, countering MCA's contractor-driven organizational barriers.

When disputes arise between contractors over interfacing DIs—such as mismatched outputs—the SDA arbitrates, using the TDP's open standards and the system level objective function in tradespace decisions to resolve conflicts impartially. As a government extension, it streamlines security by eliminating subcontractor classification chains, facilitating swift contractor replacement. The SDA sets the stage for design and production, ensuring modularity and flexibility without engaging in the detailed work itself.

Design Agents: Navigators of Subsystem Complexity

Design Agents (DAs) step into Distributed Acquisition for systems of significant scale, managing subsystems where the SDA's oversight alone proves insufficient. Each DA oversees a specific cluster of DIs—such as propulsion or electronics—translating the TDP's objective functions into area specific objective functions and detailed requirements. Rather than receiving top-down directives, DAs craft these objective functions and requirements to align with their subsystem's purpose, evaluating design choices objectively against goals like efficiency or performance. This autonomy ensures subsystems contribute to overall utility, avoiding the misaligned priorities that have challenged MCA, as noted in Lewis et al. (2025).

Within their domain, DAs arbitrate conflicts between DIs—resolving issues like incompatible specifications—using the TDP's open interfaces to adjust designs. When disputes cross subsystems, DAs collaborate with peers and escalate unresolved issues to the SDA for final arbitration. Operating as government extensions, DAs remain neutral, free from contractor incentives, and focus on coherence without designing DIs themselves. Their presence scales the methodology, distributing responsibility to manage complexity while preserving the distributed structure, ensuring large systems remain adaptable and aligned with DoD objectives.

Design Entities: The Builders of Developmental Items

Design Entities form the operational core of Distributed Acquisition, comprising a diverse group—traditional contractors, nontraditional firms, government labs, test facilities, and research centers—that bid on, design, and build DIs. They participate in the consortium to shape the reference architecture, contributing practical expertise before competing for individual DIs. Each entity works within the 150-person limit and the digital environment, leveraging the TDP's standards, tailored objective functions, and requirements to develop their assigned components. This diversity broadens participation, countering the narrow industrial base that has limited traditional MCA programs, as highlighted in Lewis et al. (2025).

The smaller scale of DIs reduces the incentive for legal challenges by losing bidders, while the TDP's openness facilitates replacement of underperforming entities. Design Entities benefit from the FFP plus military utility bonus, pushing beyond baseline requirements to earn rewards for enhanced capability. This competitive dynamic drives innovation, supported by the government-provided tools that eliminate upfront costs, ensuring even small players can contribute. Their role transforms systems into collaborative yet independent efforts, enhancing resilience and flexibility across the acquisition process.

The Inquisition Team: Guardians of Accountability

The Inquisition Team serves as an on-demand overseer in Distributed Acquisition, activated by the DoD to address delays, cost issues, or performance shortfalls, fostering shared consciousness across Lewis et al. (2025). By operating independently and reporting directly to the government, it ensures transparency and accountability, countering the hierarchical opacity that hindered oversight in MCA programs like FCS (Lewis et al., 2025). Unlike other entities, it is not a permanent fixture but a separate group reporting directly to the government, bypassing the SDA or DAs to maintain independence. It investigates root causes—whether contractor failure, oversight lapses, or unavoidable factors—recommending actions like replacing an entity or adjusting operations. This flexibility ensures accountability, addressing gaps that have persisted in MCA, as noted in Lewis et al. (2025).

When engaged, the Inquisition Team can propose significant shifts—second-sourcing a DI, adding quality controls, or even replacing the SDA—ensuring the system stays on track. Its temporary nature minimizes overhead while maximizing impact, offering a corrective mechanism that adapts to issues without entrenching bureaucracy. By standing apart, the Inquisition Team safeguards the methodology's integrity, ensuring systems deliver on their intended purpose despite setbacks.

Interactions Among Entities in Distributed Acquisition

The strength of Distributed Acquisition lies in the deliberate interactions between its entities—the SDA, DAs, Design Entities, and the Inquisition Team—structured to deliver adaptable, government-controlled systems. These relationships focus on meaningful connections that advance the methodology's goals, avoiding unnecessary overlap that could clutter the process. The SDA and Design Entities collaborate early and often, DAs mediate between the SDA and Design Entities in complex efforts, and the Inquisition Team engages all parties as an independent check when required. This selective interplay ensures a cohesive yet distributed effort, distinct from MCA's centralized hierarchies.

These interactions operate within a framework where roles remain separate—no entity can double as SDA, DA, Design Entity, manufacturer, or Integrator—preserving competition and flexibility. The following sections detail how these entities connect, illustrating a system that fosters innovation and resilience while maintaining DoD authority throughout the acquisition life cycle.

Collaborative Foundations: SDA and Design Entities in the Consortium

The SDA and Design Entities forge their partnership in the consortium, where the SDA leads a diverse group to define the reference architecture. This collaboration ensures the system's foundation reflects collective input—spanning manufacturing insights, testing needs, and technological possibilities—rather than a single contractor's agenda. The SDA synthesizes this into the TDP's conceptual level, setting objective functions for DIs that guide subsequent bids. This early engagement establishes a shared baseline, enabling Design Entities to compete on equal footing once the architecture is set.

Post-consortium, the SDA supports Design Entities by providing the digital environment and arbitrating DI conflicts. As entities design their components, the SDA ensures interface compatibility, resolving disputes—such as mismatched specifications—using the TDP's open standards. This oversight maintains system integrity without stifling contractor autonomy, while the ability to re-bid DIs keeps entities accountable. This ongoing relationship anchors Distributed Acquisition in government control, leveraging Design Entities' expertise to deliver a modular, adaptable outcome.

Bridging the Gap: DAs as Intermediaries Between SDA and Design Entities

In larger systems, DAs connect the SDA's system-level vision with the Design Entities' detailed execution, managing subsystems to scale the methodology effectively. The SDA assigns DAs clusters of DIs, handing over the TDP's developmental design with objective functions. DAs then work with Design Entities, refining these into requirements that align with subsystem goals—ensuring components contribute to overall utility. This mediation distributes oversight, preventing the SDA from becoming overwhelmed while guiding entities without dictating their designs.

DAs also resolve conflicts within their subsystems, adjusting DI designs via MOSA interfaces when issues arise. When disputes span subsystems, DAs collaborate with peers and Design Entities, escalating to the SDA if needed. This tiered approach maintains momentum, ensuring subsystem coherence integrates into the broader system. By acting as neutral intermediaries, DAs enhance the methodology's ability to handle complexity, supporting the SDA and Design Entities in delivering a unified, flexible platform.

Design Entities and Their Competitive Dynamics

Among Design Entities, interactions shift from collaboration to competition once the consortium phase concludes, a deliberate design to drive innovation and prevent consolidation. During the initial architecture definition, these entities—traditional contractors, nontraditional firms, labs, test facilities, and research centers—work together under the SDA's guidance, sharing insights to shape a system that serves all participants. Once Developmental Items (DIs) are defined and bidding begins, their relationship transforms into a competitive landscape. A contractor winning a DI is barred from bidding on adjacent or contained items, a rule enforced by the SDA to ensure no single entity dominates interconnected components, maintaining a distributed structure that avoids the monopolistic tendencies MCA has exhibited, as noted in Lewis et al. (2025).

This competitive dynamic plays out as Design Entities develop their DIs within the government-provided digital environment, adhering to the TDP's open interfaces. While they do not collaborate directly, their work intersects at these boundaries, where compatibility is critical—say, a sensor's output aligning with a processor's input. The SDA or DAs oversee these junctures, but Design Entities focus on their individual contributions, striving to maximize the system level military utility objective function established in the consortium. This system level function, central to the Value Driven Design Methodology and tied to each DI's objective function, fuels competition—entities vie to deliver enhanced capability, exercising an incentive compensation portion beyond the Firm Fixed Price if the government customer agrees to its inclusion in the design. This mechanism of option incentive pushes the DI to innovative solutions in both performance, cost, schedule, and positive life-cycle impacts where traditional approaches often stagnate. The smaller scale of DIs lowers the stakes of losing a bid, reducing legal challenges and enabling swift replacement, fostering a fluid, competitive ecosystem.

The small size of the DI also allows lower risk for concurrent development of more innovative, but potentially more risky solutions without major impact to the cost and schedule of the main effort.

The diversity of Design Entities amplifies this competition's impact. A small firm might leverage agility to innovate on a circuit card, while a lab pushes boundaries on a sensor, and a traditional contractor refines a structural component—all within the same system. This interplay, whose efforts are aligned through the overall system objective function, ensures a broad range of solutions, enhancing the system's overall quality and adaptability. The government's provision of thin clients levels the playing field, allowing entities to compete based on merit rather than resources, a contrast to MCA's bias toward established players. Through this competitive tension, Design Entities collectively build a system that benefits from varied expertise, delivering resilience and flexibility that monolithic efforts struggle to achieve.

The Inquisition Team's Oversight Role

The Inquisition Team engages with other entities in Distributed Acquisition as an independent overseer, activated only when the system encounters significant issues—delays, cost overruns, or performance shortfalls. Hired directly by the DoD, it operates outside the routine chain, reporting findings without filtering through the SDA or DAs, who might be implicated in the problem. This independence allows the team to interact with all parties—SDA, DAs, and Design Entities—to diagnose root causes, whether a contractor's failure, an oversight lapse, or an unavoidable setback. Its role ensures accountability, addressing deficiencies that have historically persisted in MCA, as outlined in Lewis et al. (2025), without embedding permanent bureaucracy into the methodology.

When engaged, the Inquisition Team's interactions with specific entities deepen based on the issue at hand. If a Design Entity struggles to deliver a DI—perhaps missing performance targets—the team investigates, recommending replacement if warranted, which the SDA executes via the TDP's open interfaces. Should the SDA falter—say, mismanaging arbitration or the digital environment—the team might propose a new SDA, a significant intervention to restore system integrity. With DAs, the team assesses subsystem oversight, suggesting adjustments if requirements misalign with objectives. These recommendations can extend to programmatic shifts—initiating a second source for a DI or implementing new quality controls—ensuring the system regains momentum without derailing the broader effort.

This targeted engagement enhances the methodology's adaptability. The Inquisition Team collaborates with the SDA and DAs to enact its proposed fixes (unless the recommendation to the government is a modification or replacement involving the SDA/DA), leveraging their authority to realign operations, while its direct access to Design Entities ensures granular insight into performance issues. Its temporary nature keeps it lean, minimizing overhead while maximizing impact, a flexibility that contrasts with MCA's entrenched oversight challenges. By interacting with all entities as needed, the Inquisition Team acts as a guardian, ensuring the system delivers on its intended purpose despite setbacks, reinforcing Distributed Acquisition's resilience.

Absence of Unnecessary Connections

Not every potential interaction among entities in Distributed Acquisition warrants emphasis, as some lack meaningful contribution to the methodology's goals. The Inquisition Team, for instance, has no routine engagement with Design Entities outside specific investigations, preserving its role as a reactive overseer rather than a constant presence. Similarly, DAs do not connect directly with the Inquisition Team unless a review necessitates it, maintaining the team's impartiality. Design Entities, despite their numbers, interact with each other only indirectly through the SDA and DAs via the TDP's interfaces, as post-consortium collaboration could undermine the competitive drive that fuels innovation.

This selectivity keeps the methodology streamlined, avoiding the cluttered relationships that can bog down large systems. The SDA and DAs focus on their respective oversight roles without redundant overlap, while Design Entities concentrate on their DIs without needing direct ties to the Inquisition Team in normal operations. By limiting interactions to those that advance modularity, control, and adaptability, Distributed Acquisition ensures a lean, purposeful structure, distinct from the hierarchical tangles that have hampered MCA. This clarity of purpose enhances the system's efficiency, directing effort where it matters most to achieve DoD objectives.

Reimagining Acquisition for the Future

Distributed Acquisition offers a bold reimagining of how the DoD acquires major capability systems, designed to deliver platforms that are modular, innovative, and adaptable to an evolving operational landscape. This methodology shifts away from MCA's centralized paradigm, actively dividing systems into Developmental Items managed by a diverse set of entities under government oversight, countering the complexity critiqued in Lewis et al. (2025). It addresses the structural issues that have long undermined acquisition efforts, as detailed in Lewis et al. (2025), providing a framework that aligns with the DoD's vision of resilient, sustainable systems. By outlining what this approach entails—its principles, entities, and their interactions—this white paper presents a transformative path forward, leaving practical execution for subsequent exploration.

The urgency of this shift stems from the DoD's need to counter advancing threats with systems that can evolve rapidly, yet transitioning to Distributed Acquisition requires addressing potential costs and contractor resistance, as noted in Lewis et al.'s (2025) discussion of post-9/11 consolidation. Phased pilot programs, such as those supported by the FORGE Act, can test the methodology on smaller systems, leveraging Value-Driven Design's cost savings to build stakeholder confidence while mitigating risks (Lewis et al., 2025).

Distributed Acquisition delivers this capability by breaking systems into smaller, manageable pieces, fostering competition, and retaining control through a government-owned TDP. The SDA, DAs, Design Entities, and Inquisition Team work together to ensure systems meet warfighting needs without the rigidity of past approaches. This methodology does not merely adjust existing processes but redefines them, offering a system that can integrate new technologies, adapt to operational shifts, and withstand disruptions—qualities essential for maintaining strategic advantage.

Restoring Clarity and Control

Distributed Acquisition actively clarifies system requirements by defining objective functions during the consortium phase, addressing the vague requirements critiqued in Lewis et al. (2025). The SDA and Design Entities collaborate to define these goals—specific, measurable targets for each DI—ensuring alignment with overall utility from the start. DAs refine these into subsystem requirements, maintaining focus on purpose rather than allowing vague specs to drift, a problem Lewis et al. (2025) identifies in MCA. The financial structure reinforces this clarity, rewarding entities that exceed baseline expected utility with an incentive reward, driving performance that directly supports DoD needs rather than minimal life-cycle and operational performance compliance at maximal cost and schedule outcomes.

Government control is reestablished through the TDP and digital engineering environment. The SDA's ownership of technical data—spanning conceptual, developmental, and production phases—enables the DoD to re-bid any DI, ensuring flexibility without dependency on a single contractor. The digital environment, managed by the government, provides tools and security via thin clients, centralizing authority and eliminating subcontractor complexities. This control allows the DoD to adjust course—replacing entities or shifting production—with minimal disruption, a stark contrast to MCA's loss of oversight, empowering the government to steer systems toward enduring effectiveness.

Fostering Modularity and Resilience

Modularity lies at the heart of Distributed Acquisition, embedded through MOSA at DI boundaries. The SDA ensures these interfaces remain open, allowing components to be swapped or upgraded without proprietary barriers, enabling systems to evolve as needs change. This approach ensures that advancements—new sensors, power units, or structural elements—can be integrated incrementally, maintaining relevance over time. The TDP's documentation supports this modularity, providing a blueprint that any contractor can use, freeing systems from the lock-in that has constrained MCA adaptability.

Resilience emerges from the methodology's distributed structure and smaller DI scale. Capping design efforts at 150 persons keeps teams agile, reducing the risk of failure cascading across the system. The ability to second-source or replace a DI—facilitated by the TDP and overseen by the SDA—minimizes disruption, while the Inquisition Team's interventions ensure rapid correction of setbacks. This resilience allows systems to withstand contractor issues or operational shifts, offering the DoD options to pursue experimental designs or dual suppliers, enhancing technological robustness and readiness for unexpected challenges.

Broadening the Defense Industrial Base

Distributed Acquisition expands the industrial base by inviting a diverse array of Design Entities—small firms, labs, and nontraditional players—into the acquisition process, echoing historical successes like the interwar period's radar development, where decentralized teams drove innovation under resource constraints. This approach, akin to WWII's network of 18,000 small firms, counters MCA's reliance on five prime contractors, fostering competition and resilience as demonstrated in distributed models (Lewis et al., 2025). The government-provided digital environment eliminates entry barriers, allowing these entities to compete on equal terms without significant upfront investment. The smaller scale of DIs reduces the stakes of bidding, discouraging legal disputes and encouraging participation from a wider pool, countering the narrow base that has limited MCA, as noted in Lewis et al. (2025). This inclusivity fosters a vibrant ecosystem where varied expertise drives system quality.

The competitive dynamics among Design Entities amplify this effect. By barring winners from adjacent DIs, the SDA ensures no single player dominates, maintaining a broad contributor base throughout design and production. The Integrator's focused role in assembly preserves this diversity, relying on inputs from multiple manufacturers rather than a single source. This approach revitalizes the DoD's supplier network, enhancing competition and innovation, ensuring systems benefit from a range of perspectives and capabilities, a strategic asset in an era of complex threats.

Aligning with DoD Objectives

Distributed Acquisition aligns seamlessly with the DoD's overarching objectives of delivering systems that embody modularity, positive life-cycle outcomes, innovation, and adaptability, goals that have proven elusive under the traditional MCA framework. Modularity is woven into the methodology's fabric through the consistent application of the Modular Open Systems Approach (MOSA) at the boundaries of each Developmental Item (DI). The System

Design Agent (SDA) enforces this standard, ensuring that every component—whether a sensor, power unit, or structural piece—connects via open, documented interfaces preserved in the TDP. This design allows the DoD to integrate new capabilities or replace outdated elements without dismantling the entire system, providing a platform that can evolve as operational requirements shift. The government's ownership of the TDP reinforces this modularity, offering a blueprint that any contractor can use to contribute, ensuring systems remain flexible and relevant over time.

Positive life-cycle outcomes emerge from the methodology's emphasis on government control and resilience throughout a system's lifespan. The SDA's retention of technical data across conceptual, developmental, and production phases empowers the DoD to manage sustainment without reliance on a single vendor, a flexibility that ensures components can be maintained, upgraded, or replaced as needed. Design Agents (DAs) contribute by aligning subsystem requirements with clear objective functions, embedding life-cycle considerations—like durability or efficiency—into the design process from the outset. The ability to re-bid DIs, supported by the TDP's open interfaces, means that if a contractor's performance wanes during sustainment, the DoD can introduce a new provider with minimal disruption. This approach delivers systems that endure, capable of supporting warfighting needs across their operational life, addressing shortcomings Lewis et al. (2025) identify in MCA's life-cycle management.

Innovation thrives within Distributed Acquisition through a structure that incentivizes and enables creative solutions. The Firm Fixed Price (FFP) paired with a military utility incentive option, calculated via consortium-defined formulas, rewards Design Entities for exceeding baseline performance—whether enhancing a sensor's range or a power unit's efficiency—pushing them to integrate cutting-edge technologies rather than settling for adequacy. The methodology's openness to a diverse pool of participants—small firms, labs, and nontraditional entities—further fuels this innovation, as varied perspectives compete to deliver superior outcomes. The government-provided digital engineering environment, accessible via thin clients, eliminates resource barriers, allowing even smaller players to propose novel approaches without the overhead that often stifles creativity in traditional acquisition. This ecosystem ensures systems benefit from the latest advancements, keeping the DoD at the forefront of technological capability.

Adaptability is a hallmark of Distributed Acquisition, enabling systems to respond swiftly to changing threats or operational demands. The smaller scale of DIs, capped at 150-person design efforts, allows the DoD to adjust individual components—swapping a sensor for a new threat profile or upgrading a structural element—without overhauling the whole platform. The SDA's arbitration role, supported by DAs in complex systems, ensures these adjustments maintain system coherence, while the Inquisition Team's oversight provides a mechanism to correct course if issues arise, such as second-sourcing a DI or shifting production priorities. This granularity and flexibility mean systems can evolve incrementally, integrating new technologies or responding to battlefield shifts, a capability MCA struggles to achieve, as noted in Lewis et al. (2025). Together, these elements deliver platforms that meet the DoD's need for agility, ensuring readiness in an unpredictable landscape.

Advancing the FORGE Act with the Distributed Acquisition Approach

The Fostering Reform and Government Efficiency in Defense Act (FORGE Act, S. 5618), introduced in December 2024, aims to streamline DoD acquisition, promote commercial contracting, and diversify the Defense Industrial Base (DIB) (Wicker, 2024). The distributed acquisition approach, by fractionating systems into Developmental Items (DIs), capping teams at 150 persons, and retaining government control via a TDP, offers a methodology to implement the FORGE Act's vision, addressing Major Capability Acquisition (MCA) flaws while fostering modularity and innovation.

The FORGE Act's push to reduce bureaucratic barriers aligns with the distributed acquisition approach's iterative bidding process, which replaces MCA's linear framework with a flexible "bid→architect→bid→design→bid→build" sequence. This minimizes delays, as seen in MCA's 2-year average overruns, by fostering competition and reducing contractor lock-in. Small teams, informed by Dunbar's 150-person coordination limit, accelerate decisions, supporting the Act's call for agility (Lewis et al., 2025). For instance, a FORGE Act-funded autonomous vehicle program could leverage Distributed Acquisition to engage small tech firms and labs as Design Entities, developing modular components like sensors and propulsion systems linked by MOSA interfaces, ensuring rapid delivery and adaptability. This approach mirrors WWII's diverse network of 18,000 firms, over half small businesses, which delivered modular systems swiftly, operationalizing the FORGE Act's push for DIB diversification and commercial contracting while countering MCA's prime contractor dominance (Lewis et al., 2025).

The Act's emphasis on commercial contracting is enabled by the approach's open TDP interfaces, allowing nontraditional firms to bid without proprietary barriers. The government's digital engineering environment lowers entry costs, inviting commercial players. Value-Driven Design (VDD) optimizes DI trade-offs, potentially saving $55 billion annually (Lewis et al., 2025), aligning with the Act's commercial focus. Diversifying the DIB, a core FORGE Act goal, is achieved by engaging small firms and nontraditional contractors, countering MCA's five-prime dominance (DoD, 2022a). A hypothetical hypersonic missile program could leverage diverse DI developers, enhancing competition and resilience, as WWII's small-firm network demonstrated (Lewis et al., 2025).

The distributed acquisition approach supports the FORGE Act's portfolio-centric acquisition by enabling Program Acquisition Executives to manage DIs holistically, aligning with McChrystal's shared consciousness model. Pilot programs, such as autonomous vehicles, could test this under the Act's flexible funding, with the Inquisition Team ensuring accountability. By operationalizing the FORGE Act, the distributed acquisition approach delivers modular, innovative systems, breaking MCA's cycle of inefficiency and ensuring warfighter readiness.

A Call to Action

Distributed Acquisition is not merely a theoretical exercise but a pressing call to action for the DoD, a response to the urgent need to break free from an acquisition model that falters under modern pressures. The methodology offers a system where the SDA, DAs, Design Entities, and Inquisition Team collaborate to produce platforms that are modular, sustainable, and innovative—qualities essential as adversaries deploy increasingly sophisticated capabilities. Lewis et al. (2025) underscore the cost of inaction: a traditional approach that embeds complexity, does not effectively communicate military utility, cedes control, and limits adaptability, costing the DoD $208 million daily in losses, causes strategic lag against more agile adversaries.

Distributed Acquisition counters this by fractionating systems into manageable pieces, broadening the industrial base, and retaining government authority, delivering a framework that aligns with the DoD's mission to maintain enduring advantage.

The stakes demand bold change. Systems acquired through this methodology can integrate new sensors, propulsion, or electronics as threats evolve, supported by a TDP that ensures flexibility without proprietary constraints. The diverse participation of Design Entities—enabled by a government-managed digital environment—revitalizes the DoD's supplier network, fostering competition and innovation where MCA narrows options. The SDA's oversight, bolstered by DAs and the Inquisition Team, ensures these systems remain on track, adaptable to setbacks or shifts in priority, offering resilience that traditional large-scale contracts lack. The result is a vision of acquisition that delivers platforms ready for today's fight and tomorrow's challenges, a necessity as operational tempos accelerate.

Acquisitions professionals must seize this opportunity to redefine how the DoD builds its capabilities. The methodology's distributed structure, with its emphasis on modularity and control, provides a system that can pivot—second-sourcing a component, replacing an underperformer, or pursuing experimental designs—without the inertia that has bogged down past efforts. The smaller scale of DIs reduces risk, enabling the DoD to test and refine technologies incrementally, while the integrator's focused role ensures these pieces coalesce into a unified whole. This approach does not require abandoning all past lessons but builds on them, offering a practical path to systems that meet warfighting needs with agility and precision.

This approach addresses concerns and impediments to efficient acquisition that have remained persistent through 40+ years of reform; the DoD needs a methodology that delivers results now and adapts for the future. Distributed Acquisition provides that framework, a system where government-led architecture, competitive design, and independent oversight converge to produce platforms that endure. It empowers the DoD to harness a broad industrial base, integrating diverse expertise into systems that can shift with the threat landscape—whether countering new technologies or sustaining operations over decades. This white paper defines the "what": a transformative approach that restores clarity, fosters resilience, and aligns with strategic goals, urging the DoD to act decisively to implement this vision.

The call extends beyond process to purpose. Distributed Acquisition ensures the DoD can field systems that keep pace with adversaries, delivering modularity that allows upgrades, innovation that pushes boundaries, and adaptability that meets emerging needs. The SDA anchors this effort with a vendor-agnostic foundation, DAs scale it to complex systems, Design Entities build it with diverse input, and the Inquisition Team safeguards its integrity. Together, they offer a system that reclaims control from centralized pitfalls, as highlighted in Lewis et al. (2025), providing the DoD with platforms that prevail in an era of relentless change. This is a potential future of acquisition where a distributed, dynamic approach amends a few of the unchanged elements The acquisition organization is responsible for resolving recurring and negative development outcomes to secure the DoD's mission for future generations.

References

Lewis, R. P., Carnes, J., & Thompson, J. (in press, 2025). Finding opportunities in the adaptive acquisition framework: A concept for a distributed acquisition approach. In *Creating synergy for informed change: Transitioning technology to the warfighter, NPS 22nd Annual Acquisition Research Symposium and Innovation Summit.*

THIS PAGE INTENTIONALLY LEFT BLANK

Applying Agile to Mitigate the Risk of Transitioning Defense Embedded Software Technologies

Salvatore (Sam) Fazio—teaches acquisition and program management, test and evaluation, leadership, and engineering and technical management classes to the Department of Defense (DoD) acquisition workforce in his current position at the Defense Acquisition University (DAU). He also has a Bachelor of Science in Astronautical Engineering from the United States Air Force Academy and a Master of Science in Aeronautical and Astronautical Engineering from Purdue University where he studied optimization of electrical and mechanical control systems. [salvatore.fazio@dau.edu]

Karen LaFond—teaches engineering and software program management courses to the DoD acquisition workforce in her current position at the Defense Acquisition University. Karen is a founding member of the DAU SW Cadre, which is designed to promote increased collaboration and knowledge sharing among DAU business units and external stakeholders, specifically between the DAU and OUSD(A&S) Software Cadres. Karen has a Bachelor of Science in Computer Science from Michigan State University, and a Master of Science in Software Engineering and a Master of Science in Program Management, both from the Naval Postgraduate School. [karen.lafond@dau.edu]

Abstract

The Department of Defense utilizes emerging embedded software technologies to enhance its warfighting capabilities; however, the transition process is often inefficient and ineffective, and carries certain risks. Weapon system development typically occurs within a traditional bureaucratic framework characterized by heavy regulatory restrictions, a culture laden with security constraints, hierarchical decision-making, funding limitations, safety concerns, lack of adherence to processes, communication/coordination failures, and risk averse mindsets. Promising technologies often face significant obstacles, which can hinder or prevent their progression from development to operational use. Applying modern software acquisition and development principles to the embedded software technology transition processes can boost efficiency and effectiveness, and mitigate risk. As a result, the technology that our warfighters need could be delivered "at the speed of relevance" (Defense Innovation Board, 2019, p. 2).

This paper will identify which of the 12 Agile Manifesto principles are effective in reducing inefficiencies when applied to defense programs, and will outline the benefits that can be realized. Additionally, it will provide examples of successful implementation demonstrated by current and past defense programs. In conclusion, while it is an investment that will take resources and time before realizing results, we assert that it will be well worth it to improve the likelihood of success to transfer technology for use on modern defense systems.

Keywords: Embedded Software, Transition, Risk, Agile, Valley of Death

Introduction

The world has witnessed a rapid increase in new technologies. Technologies such as generative AI, quantum and edge computing, autonomous vehicles and drones, cybersecurity, augmented reality, and robotics have enabled unprecedented capabilities and new opportunities. The United States Department of Defense (DoD) has advanced new technologies to defensive and offensive systems making military operations more efficient, effective, and safer for personnel. DoD however is encountering many challenges with incorporating new technologies. Integration, to ensure compatibility and interoperability between old and new systems, is one of these challenges facing the DoD. Others include cybersecurity risks, training and skills gaps, cost and budget constraints, and regulatory and ethical concerns.

This paper investigates the benefits and opportunities of using modern agile principles to significantly aid in transitioning new technologies. The premise is that the transition of technology within an agile environment provides the relevant backdrop to mitigate or significantly reduce the obstacles facing current DoD technology transfers. Other benefits of establishing an agile setting include an increase in flexibility and responsiveness, reduced program risks, enhanced collaboration, and continuous improvement, all of which facilitate DoD technology transfer.

We researched a variety of DoD programs intended to integrate new technologies to meet growing performance requirements or fill a capability gap. We found that these programs were influenced by typical forces, which could have restrained their success. Our investigation revealed that successful programs implemented the agile principles with a growth mindset.

Our first task was to identify those forces which increased the likelihood of failure (restraining forces) and those which propelled the program to success (driving forces). By reviewing program historical files, we discovered that most acquisition programs experience a similar and recurring set of restraining forces due to the traditional culture and the environment of DoD program offices. Without intervention, and if left unchecked, these acquisition programs would likely have been unsuccessful. Consequently, the list of restraining forces, as shown in the Force Field Analysis (Figure 1), frequently impact DoD systems. On the other hand, the driving forces are characteristic of highly successful programs which implement the 12 principles identified in the Agile Manifesto.

Agile Methodology and Principles

Agile methodologies emphasize iterative development, which allows for continuous feedback and improvements. This is particularly useful when transitioning technology, as it enables teams to adapt to changes and address issues promptly. Agile and DevOps methodologies contribute to customer-centric approaches and service optimization. By integrating these methodologies, organizations can foster a culture of innovation and continuous improvement. Agile is quickly overtaking waterfall as the methodology for developing products.

Making the switch from “traditional” project management to agile is not always straightforward, and it can be particularly challenging for organizations that are accustomed to a predictive environment. However, with the right guidance and support, the transition doesn't have to be overwhelming.

Agile technology supports innovation by fostering a flexible and adaptive environment that encourages continuous improvement and rapid iteration. Here are some key points on how agile methodologies contribute to innovation.

1. **Customer Satisfaction**: Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.
2. **Welcome Change**: Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.
3. **Frequent Delivery**: Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.
4. **Collaboration**: Business people and developers must collaborate daily throughout the project.
5. **Motivated Individuals**: Build projects around motivated individuals. Give them the environment and support they need and trust them to get the job done.
6. **Face-to-Face Conversation**: The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.
7. **Working Software**: Working software is the primary measure of progress.

8. **Sustainable Development**: Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
9. **Technical Excellence**: Continuous attention to technical excellence and good design enhances agility.
10. **Simplicity**: Simplicity—the art of maximizing the amount of work not done—is essential.
11. **Self-Organizing Teams**: The best architectures, requirements, and designs emerge from self-organizing teams.
12. **Retrospection**: At regular intervals, the team reflects on how to become more efficient, then tunes and adjusts its behavior accordingly.

Using Force Field Analysis to Drive Change

In this paper we apply the Force Field Analysis management tool to diagnosis the internal and external forces that form the DoD, program office, acquisition environment and culture. With the forces in balance and at equilibrium, the environment and culture remain stable. To change the environment, forces, whether driving (positive) or restraining (negative), need to be either strengthened or weakened. The driving (positive) forces are defined as those which will produce a change in the organization toward the desired direction. For example, an organization may want to establish a culture that fosters open communications. Any force which will move the organization in that direction is considered a driving or positive force. Restraining (negative) forces, on the other hand, are those forces that will result in producing an environment opposite to what is desired. The intent is to weaken the restraining forces so that the driving forces overcome those restraining forces and generate the desired outcome, namely, an acquisition environment and culture that support the application of agile principles.

In this paper, the authors selected restraining or negative forces based on the information from program historical files. For example, if a program were plagued by a backlog of volatile requirements and unstable priorities reducing progress or causing the program to stagnate, volatile requirements and unstable priorities was selected as a restraining force.

The purpose of using the Force Field Analysis is to change the program office culture so that as resources (financial, manpower, schedule, material, etc.) are applied, the acquisition program avoids stagnation and quickly progresses through the required acquisition phases. Stagnation becomes a roadblock and prevents a program from advancing to successful completion. This stagnation is often termed the “Valley of Death.” Applying resources with no effect eventually causes a program to languish and likely “die” or be canceled.

We discovered that this acquisition gap or period of stagnation typically lasts for 1–2 years, during which program resources of adequate finances, available time and manpower, and acceptable materials, fail to move the program forward, resulting in the program dying in the proverbial “Valley of Death.”

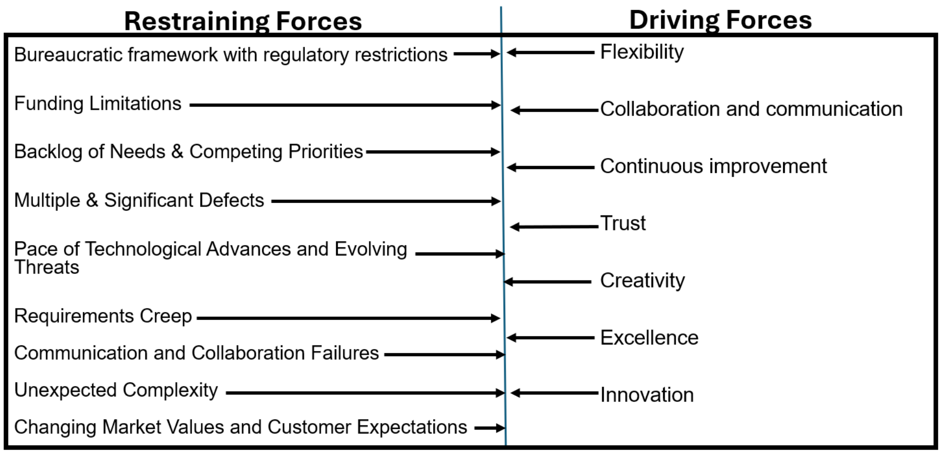


Figure 1. Force Field Analysis for Implementing an Agile Environment

Sample Programs Implementation

Each of the following defense programs demonstrated an improved acquisition outcome by mitigating one or more of their identified challenges, some of which were restraining forces considered for this paper.

Restraining Force – Bureaucratic Framework & Restrictions

Program – Army Robotic Combat Vehicle

Proposed Application of Related Agile Principles

Highlighted principles indicate author's interpretation of which were applied on the Army Robotic Combat Vehicle program.

Agile Principles	1 - Customer Satisfaction	2 – Welcome Change	3 – Frequent Delivery	4 –Collaboration	5 – Motivated Individuals	6 – Face to face Conversation	7 – Working Software	8 – Sustainable Development	9 – Technical Excellence	10 - Simplicity	11 – Self Organizing Teams	12 - Retrospection

Program Analysis

The program office for the Robotic Combat Vehicle (RCV) Program began during DoD's full endorsement and implementation of its traditional, heavily burdened, acquisition framework. This framework is overwhelmed with regulatory restrictions and scrupulous compliance. The entire acquisition process was laden with required documented processes and overbearing oversight. The objectives of the RCV program were to achieve advanced system performance and to integrate innovative technologies such as artificial intelligence for autonomy and self-decision-making. Consequently, technical requirements were highly volatile and often delayed progress on the program even though substantial resources were allocated and consumed. Frequently, the decision-makers in the program office met with significant bureaucratic resistance to implement a change in the traditional process.

When a change in the acquisition process was accepted and the agile principles were incorporated, the program experienced several characteristics commonly found in a program conducted within an agile environment. When the program office implemented the Middle-Tier of Acquisition (MTA) pathway concurrently with the Software Acquisition (SWA) pathway as described in the Adaptive Acquisition Framework (AAF), delivery time for the individual embedded software increments significantly decreased, while capability releases provided an increase in performance from the technologies that were promptly integrated into the release. The new embedded technologies enabled the integration of artificial intelligence in the RCV, which permits the processing of vast amounts of data and informed decision-making in real time situations. The operational units receiving the rapid deployment of upgraded robotic systems also noted an increase in operational efficiency and effectiveness.

These technologies which prototyped autonomous software and processes for the RCV program received increased prioritization in the development effort due to rapid prototyping and the flexibility to “fail fast, fail often, and learn.” As learning took place, the program remained adaptable to accommodating the constant adjustments and evolving technical requirements necessary to meet the changing needs of the program and to leverage industry experience and expertise.

Not only did the agile environment produce significant benefits for the RCV system, but it also supported an improvement in the acquisition culture and in the way the program office was managed. The program office received more frequent and consistent input from department leadership with reduced bureaucratic hurdles and barriers. With the reduced administrative framework and regulatory restrictions, monthly governance meetings were conducted along with Component Acquisition Executive (CAE) quarterly meetings to discuss the MTA portfolio. These meetings helped to streamline decision-making and further reduce managerial delays.

Restraining Force – Funding Limitations

Program - VH-92A Presidential Helicopter

Proposed Application of Related Agile Principles

Highlighted principles indicate author’s interpretation of which were applied on the VH-92A Presidential Helicopter program.

Agile Principles	1 - Customer Satisfaction	2 – Welcome Change	3 – Frequent Delivery	4 –Collaboration	5 – Motivated Individuals	6 – Face to face Conversation	7 – Working Software	8 – Sustainable Development	9 – Technical Excellence	10 - Simplicity	11 – Self Organizing Teams	12 - Retrospection

Program Analysis

By implementing agile principles, the VH-92A Presidential Helicopter program office overcame significant funding constraints and ultimately reduced the cost of the program by about 10% of its original 2014 estimate, while still achieving program objectives and meeting system requirements.

The defense contractor, Sikorsky, a Lockheed Martin Company, worked together with the program office to achieve substantial cost savings. The program office developed an adaptable acquisition strategy, which Sikorsky executed. Through collaboration between the contractor and the government technical workforce, Sikorsky was able to deliver technically excellent products. The Navy program office worked effectively with its contractors to address any technical and management issues and to implement coordinated solutions. The acquisition strategy identified the Navy program office as the Lead System Integrator (LSI). In the LSI model, the Navy takes a more active role in managing the integration of developed systems into the aircraft.

The decision to begin with and modify the FAA-certified Sikorsky, S-92 aircraft, known for its safety and reliability, produced perhaps the most substantial reduction in development costs. As a baseline prototype system, the program office integrated government-defined mission systems and high-tech capabilities into the commercial helicopter. As a result, the contractor was able to tailor and streamline its processes for increased efficiency. As a priority, Sikorsky developed working-software, which provided extensive secure and non-secure communication systems. The Navy worked closely with Sikorsky and other contractors to include electromagnetic pulse hardening and a crash-survivable, flight-information recorder sensor. In addition, the government-contractor team resolved issues related to electromagnetic event survivability and landing zone suitability. Technical features such as day/night and all-weather operations, a self-contained navigation system, and a global positioning system were included to deliver a fully high-tech, capable system.

The program office took advantage of additional cost saving initiatives. For example, instead of contracting for maintenance, the Navy leveraged its existing resources of personnel and facilities, as well as its established infrastructure. The effective use of available resources and infrastructure avoided the need to send the aircraft back to the developer for maintenance. Another example of implementing a substantial, cost-saving, initiative relates to part shortages. The lack or delay of aircraft maintenance parts increases costs and results in significant flight downtime. To resolve this cost-related issue, the program office initiated a tracking system to identify and resolve the delay of parts.

Restraining Force – Backlog of Needs & Competing Priorities

Program - Joint Program Executive Office - Chemical, Biological, Radiological, and Nuclear Defense

Proposed Application of Related Agile Principles

Highlighted principles indicate author's interpretation of which were applied on the Chemical, Biological, Radiological, and Nuclear Defense (CBRND) program.

Agile Principles	1 - Customer Satisfaction	2 – Welcome Change	3 – Frequent Delivery	4 –Collaboration	5 – Motivated Individuals	6 – Face to face Conversation	7 – Working Software	8 – Sustainable Development	9 – Technical Excellence	10 - Simplicity	11 – Self Organizing Teams	12 - Retrospection

Program Analysis

Perhaps one of the most challenging tasks in weapon system development is to establish operational requirements for a robust and effective warfighting capability. As challenging as this is, the effort becomes even more overwhelming when considering multiple factors such as the operational environment, the cultural background of stakeholders, organizational responsibilities, tight schedules, etc. Each of these interests often have a set of volatile and unstable requirements which are unaligned and compete with others for consideration. This is the situation in the CBRND program office, which is charged with informed decision-making in support of competing and often conflicting joint force requirements. The goal of the CBRND program is to integrate advanced sensors and data analytic technologies for medical countermeasures, protective equipment, and detection systems into existing systems and platforms to enhance threat detection and response.

The design and improvement of chemical and biological detection technologies are frequently occurring and provide advanced capabilities. However, the selection of which capabilities to acquire or integrate depends on competing priorities, stakeholders' operational needs, and mission objectives. The amount of cutting-edge technology insertion directly impacts the deployment schedule. While both the need for technology and field-deployment is driven by operational and possibly combat conditions, the joint program office needs to provide a balanced resolution.

Developing new technologies comes at a cost, and decisions need to be made between maintaining current capabilities and acquiring innovative, chemical and biological detection systems. Funding for research programs is often limited due to the accelerated pace of technology development and the scarcity of designated funds for research. Likewise funding for maintenance of fielded systems is competitive. Therefore, the program office needs to prioritize funding priorities not only among promising technologies, but also between maintaining the capability of current and legacy systems and acquiring novel and advanced technologies.

The program office applied several agile principles leading to successfully achieving program objectives. For example, when capability demonstrations and engagements were conducted or when decisions on platforms were considered, all branches of service and stakeholders were invited to observe so that quick pivots and reprioritization of user needs were made from the feedback of all interested participants.

Restraining Force – Multiple & Significant Software Defects

Program – ARMY PEO Enterprise Information Systems

Proposed Application of Related Agile Principles

Highlighted principles indicate author's interpretation of which were applied on the Army Program Executive Office (PEO) Enterprise Information Systems program.

Agile Principles	1 - Customer Satisfaction	2 – Welcome Change	3 – Frequent Delivery	4 –Collaboration	5 – Motivated Individuals	6 – Face to face Conversation	7 – Working Software	8 – Sustainable Development	9 – Technical Excellence	10 - Simplicity	11 – Self Organizing Teams	12 - Retrospection

Program Analysis

The vision of the Army's PEO Enterprise effort is to transform the Army's current business acquisition organization into an agile environment taking advantage of speed, simplicity, and excellence in delivering capabilities. To fulfill this vision, PEO Enterprise supports every domain, branch, and unit worldwide with the expectation of modernizing, enhancing, and managing the Army's enterprise business systems that keep the Army operating efficiently.

Enterprise business systems rely on high-quality software and software products. Producing and delivering working software, that is, software free from defects, is essential to successful mission performance. Therefore, an important goal of PEO Enterprise is to reduce or eliminate significant software errors and defects. PEO Enterprise attempts to eliminate multiple and significant software defects by implementing key test strategies through a comprehensive agile transformation process. Test strategies are frequently incorporated into the software development process enabling developers and testers to quickly find and resolve software defects. Through early and continuous testing, defects are identified and resolved before becoming deeply embedded in the software design at a higher architectural level. Regular and incremental testing of software encourages frequent and continuous user feedback and assists in identifying software design criteria. The U.S. Army Test and Evaluation Command (ATEC) can significantly contribute to the process of producing error-free software by observing how software tests are designed and conducted. ATEC's review of test data during the development process adds credibility to the test evaluation and results and provides feedback to developers and other stakeholders. In addition to developmental test results, demonstrations of operational capabilities, while conducting realistic scenarios, provide necessary feedback to stakeholders, encourage collaboration, and welcome any changes required to completely satisfy the customer.

The agile practice of retrospection provides time for reflection on how well the organization is working to reduce or eliminate software defects and what changes need to be made. The retrospective considers any functional area or domain, which can help achieve its goal. For example, it may consider what further amount and type of manpower are required to effectively and efficiently test software during development, deployment, and delivery.

A retrospective considers organizational change in other functional domains which support an agile environment. These changes can also help to reduce software defects, increase delivery speed, and incorporate technical excellence. Software data, which is visible, accessible, understandable, linked, trusted, interoperable, and secure confirms the validity of accurate test results. Highly trained personnel and those skilled in developing and understanding software applications and information technology services determine the approach to developing appropriate tests and evaluate collected data. Appropriate funding and type of contract ensure adequate financial support and agreements necessary to fully implement the required software testing and review of the results. And finally, the appropriate acquisition professionals, domain specialists, needed facilities and supplies, and force structure secure the environment to produce working software free from defects and error.

Restraining Force – Slow Pace of Technological Advances & the Evolving Threat

Program - ARMY Joint Common Access Platform (JCAP)

Proposed Application of Related Agile Principles

Highlighted principles indicate author's interpretation of which were applied on the Army Joint Common Access Platform (JCAP) program.

Agile Principles	1 - Customer Satisfaction	2 – Welcome Change	3 – Frequent Delivery	4 –Collaboration	5 – Motivated Individuals	6 – Face to face Conversation	7 – Working Software	8 – Sustainable Development	9 – Technical Excellence	10 - Simplicity	11 – Self Organizing Teams	12 - Retrospection

Program Analysis

One of the restraining forces directly associated with transitioning embedded defense technologies is related to the pace at which the technology software is developed and deployed. These defense technologies are effective in countering the evolving threat. According to the 2025 Annual Threat Assessment, both the number of threats and their source are growing at an escalating rate. New advanced threat technologies, digital warfare, and system vulnerabilities are the main causes of the increase in attacks.

The Army's JCAP supports U.S. Cyber Command (USCYBERCOM) by providing a protected environment from which to execute coordinated cyber-attacks against approved targets. JCAP enables DoD cyber operators to connect to their targets and neutralize the rapidly evolving foreign threats from nations such as China, Russia, Iran, and North Korea.

To keep pace with the evolving multi-national threat, the JCAP program office must implement an agile environment for software development. Program success is measured by the effectiveness of U.S. offensive attacks against threats and the U.S. ability to defend its national infrastructure and defense systems. Success will require a software development environment which welcomes changing requirements, frequently delivers offensive and defensive cyber software, and incorporates excellence in developing cutting edge technologies. This type of environment is characteristic of an agile organization which routinely generates working software with minimal or no significant defects.

Even though the JCAP program has shown continuing success, it faces significant and ongoing challenges. JCAP integrates cyber capabilities from among several stakeholders and across defense components. Each of these capabilities are often unique in their architecture and design. JCAP is responsible for integrating these capabilities into a single platform and providing seamless operations. While maintaining operational effectiveness, JCAP is charged with incorporating advanced and emerging technologies. Even though JCAP is an offensive system, it must also provide defensive security against cyber-attacks against itself. Balancing the need for rapidly incorporating offensive and defensive capabilities and the need for thorough and rigorous testing is another challenge that requires continuous and regular assessments of the gaps that exist among threats, system requirements, available, and advance technologies.

Restraining Force – Requirements Creep

Program - Special Operations Command (SOCOM) MQ-9 Reaper Unmanned Air Vehicle (UAV)

Proposed Application of Related Agile Principles

Highlighted principles indicate author's interpretation of which were applied on the MQ-9 Reaper Unmanned Air Vehicle (UAV) program.

Agile Principles	1 - Customer Satisfaction	2 – Welcome Change	3 – Frequent Delivery	4 –Collaboration	5 – Motivated Individuals	6 – Face to face Conversation	7 – Working Software	8 – Sustainable Development	9 – Technical Excellence	10 - Simplicity	11 – Self Organizing Teams	12 - Retrospection

Program Analysis

Today's agile software development environment consists of a mix of several methods that were invented in the early 2000s. Popularity advanced quickly in private industry, but not as quickly on government software development projects. DoD issued policy and guidance in 2020; however, the new policy and guidance had not been fully adopted by most DoD programs. Since there is no one right way for every program implementation, a wide range of lessons learned are driving best practices.

Software development on the SOCOM MQ-1 and MQ-9 Combat Air Patrol programs has provided direct support to the weapon systems operational flight program for many years. When urgent requirements are received from SOCOM, the standard procedure has been to simply add capability to an existing legacy system without regard for efficiency. Because the program office relied on a software "waterfall" development, the software took time to produce. Quick and frequent releases were unheard of. Instead, a validated set of requirements and a designated verification test plan were generated in the early planning stages. If requirements changed, the contract and all documentation had to be modified. Cost was incurred and schedules suffered, since neither cost nor schedule was planned for.

So that the program office would be able to adequately respond to SOCOM's need for capability additions, the program manager, responsible for the MQ-1 and MQ-9 software development, implemented a service type contract. The contract required delivery of a repeatable process, which included 6 months of development followed by 6 months of test. It was their way to implement schedule-driven capability development. Unfortunately, they did not account for the time that would be needed to integrate and perform lab testing.

Since multiple end-user demonstrations were conducted to curb requirements creep, the program office was forced to reduce time for coding and testing. As a result, numerous critical defects were discovered, which required further rework and additional deliveries to correct the software errors and deficiencies. Consequently, previously scheduled, major software deliveries releases were delayed. Therefore, the program experienced overall significant schedule delays when software tests were not adequately conducted.

With each new scope change release, the development contractor was given a choice to either create a new Minimum Viable Capability Release (MVCR), include the new scope in the next planned Minimum Viable Product (MVP), or add the change to a later release. The software development team delivered 50+ new end-user requirements, 500+ software changes initiated by the contractor, and 100+ development and operational test defects, yet still fielded six operational flight programs. Progress was considered successful even though releases were not delivered as planned. Improvement was evident when schedule slips were reported in days and weeks instead of historically in months and years. Because end-users trusted the team to incorporate operational feedback in the next release, they were satisfied, more so than before. End-users could expect and depend on a 6-to-12-month product delivery schedule. Despite the recognized improvements in schedule, performance goals were not being met.

The development team, PMO, and end users met weekly to resolve process issues. While trusting in the current cadence, SOCOM required more performance capability, faster deliveries, and more reliable software. Consequently, the program office implemented Agile principles, for example, responding to change with more frequent deliveries of working (demonstratable) software. They increased their collaboration with end-users and emphasized the necessity for motivated individuals and frequent communications. As the software development team embraced the agile principles, they were able to increase delivery frequency, which improved customer satisfaction.

The program manager switched to using a level-of-effort type contract to avoid the complexity and time it took to make requirement changes through contract modifications. A dedicated contract team with a repeatable process was put in place to help stabilize funding. When faced with new organizational, cultural, and legal challenges, the program office developed engagement rules to control requirements volatility. They learned that an increase in the frequency and number of deliveries translated into a heavier workload for the program office in the form of oversight.

In their attempt to mimic private industry's speed to market, it became clear that military systems are more complicated and unique than commercial systems. Independent military test organizations needed for achieving system certifications were continually challenged by the uptick in delivery cadence.

Program offices must find ways to implement agile principles that may be different than how other organizations implement them. The MQ-9 Reaper program office embraced the challenge and created success. The most important lesson learned was that all team members and stakeholders had to equally commit and participate in making the changes happen.

Restraining Force – Communication & Collaboration Failures

Program - Air Force F-22 Raptor and F-35 Lightning II

Proposed Application of Related Agile Principles

Highlighted principles indicate author's interpretation of which were applied on the Air Force F-22 Raptor and F-35 Lightning II programs.

Agile Principles	1 - Customer Satisfaction	2 – Welcome Change	3 – Frequent Delivery	4 –Collaboration	5 – Motivated Individuals	6 – Face to fac e Conversation	7 – Working Software	8 – Sustainable Development	9 – Technical Excellence	10 - Simplicity	11 – Self Organizing Teams	12 - Retrospection

Program Analysis

Despite endorsement by President Trump's Executive Orders, the FY24 NDAA, the Department of Government Efficiency (DOGE) initiatives to improve efficiency and effectiveness of software modernization methods and DoD acquisition outcomes, significant challenges still exist. Long development cycles are supported by deeply entrenched habits using the Waterfall development method, while culture typically does not support modern practices. Program offices are still highly resistant to change. Implementing change requires more attention to communication and collaboration, as well as gaining the knowledge and implementation skillsets needed.

The F-22 program faced several challenges prior to modernizing efforts. Not adapting an agile mindset, vendors using traditional methods yet failing to meet schedule or performance goals, and proliferation of single, large software deliveries plagued by an overburden amount of documentation resulted in duplication of effort and test challenges, which failed business plans. Change would decrease software deliveries but was needed to move forward.

The F-35 program was also behind in capability upgrades and, like the F-22 contract, development teams did not meet performance goals. The program office faced emerging requirements, which it was not able to manage, causing schedules to slip and threatening overall cost, schedule, and performance goals. Software vendors were unable to deliver timely software updates, especially when faced with unexpected complexity. The F-35 program office began to recognize the need for standardization in the battlefield. Unfortunately, existing software development processes prevented them from achieving planned improvements.

Even with benefits from software acquisition and development policy, guidance, and best practices challenges continue to surface. Even though end-users received working software faster and more frequently, they adjusted to receiving software that did not contain the full set of requirements. There is still major resistance to change, which may be rooted in the lack of knowledge and/or understanding of modern concepts, terminology, and basic implementation methods. Consideration needs to be given when it does not make sense to apply modernization to legacy programs and when modernization would misalign program goals with leadership objectives. The tendency to continue large programs even after significant performance and cost overruns occur should not prevent more aggressive change management.

To eliminate costly changes, the DoD should determine whether software modernization decreases cost, satisfies customers, and improves product quality; but at the same time fund software modernization practices expecting a return on the investment.

Restraining Force – Unexpected Complexity

Program - Joint Tactical Radio System (JTRS)

Proposed Application of Related Agile Principles

Highlighted principles indicate author's interpretation of which were applied on the Joint Tactical Radio System (JTRS) program.

Agile Principles	1 - Customer Satisfaction	2 – Welcome Change	3 – Frequent Delivery	4 –Collaboration	5 – Motivated Individuals	6 – Face to fac e Conversation	7 – Working Software	8 – Sustainable Development	9 – Technical Excellence	10 - Simplicity	11 – Self Organizing Teams	12 - Retrospection

Program Analysis

In 1997, DoD initiated the Joint Tactical Radio System (JTRS) communication network program to significantly expand capability. Unfortunately, the program experienced constant delays due to unexpected technical complexity resulting in major cost overruns. The original plan was to implement an open architecture promising to reduce the need for future development, integration, and testing; however, it did not turn out as promised. The program office realized that the additional technology was available and ready to be added to the system; however, the program experienced inadequate integration planning.

Under joint management the program office experienced a significant lack of communication and collaboration. Program costs became uncontrollable from the many technical problems; and after system development struggled, the value of applying basic systems engineering principles became clear.

The program office was able to leverage valuable technology from the competitive environment; however, there was no way to prioritize the large number of initial requirements and interface standards. Consequently, the program struggled with managing concurrent hardware and software development.

Although there were a few early signs of adopting basic agile principles, the JTRS program remains an example of reducing unfavorable effects through an aggressive implementation of agile best practices. Unfortunately, modern software development policy was not codified and released until 2020, even though many of the agile principles were practiced in the early 2000s.

Restraining Force – Changing Market Demands

Program - John Deere

Proposed Application of Related Agile Principles

Highlighted principles indicate author's interpretation of which were applied on John Deere programs.

Agile Principles	1 - Customer Satisfaction	2 – Welcome Change	3 – Frequent Delivery	4 –Collaboration	5 – Motivated Individuals	6 – Face to face Conversation	7 – Working Software	8 – Sustainable Development	9 – Technical Excellence	10 - Simplicity	11 – Self Organizing Teams	12 - Retrospection

Program Analysis

John Deere programs demonstrate how they embraced change to the modern software development environment in keeping up with evolving technology complexities, emerging markets, customer needs, and a desire to improve the quality of software products. Use of the Waterfall software development method continued to impose longer than planned development cycles, delayed response, and/or inability to meet emerging customer demands. Company culture was tuned to long term habits that did not serve modern needs.

The company began to learn and implement SCRUM development methods, with a lean twist to facilitate short continuous development, integration, and delivery cycles. They learned and lived migration to an agile mindset while at the same time receiving training and coaching to stay on track.

John Deere management demonstrated their commitment to leading change by showing examples. They were successful at shortening their software development cycles, which enabled them to welcome change to meet emerging customer needs. Their customers became more satisfied with more and faster software deliveries, not to mention the improved product quality. They attribute their accomplishments to all the work they did to make their process more efficient and effective. Multi-discipline communication and collaboration across the company grew naturally as team members became more independent and enthusiastic.

The company learned along the way that management commitment and participation are essential, along with the willingness by everyone in the company to persist until they achieve positive results. They now understand that product and process improvement requires the continuous attention of everyone in the company and to all types of effort presented (i.e., training, communication, collaboration, etc.). They should be proud of their success!

Conclusions

Each of the examples presented in this paper applied agile principles and demonstrated an increased likelihood of successfully transitioning state-of-the-art technologies. Historical demonstration of success along with stronger directives, policy, and guidance will enable the momentum needed to accelerate changes to get capability to warfighters faster.

President Trump's executive orders related to software acquisition focus on enhancing the efficiency and effectiveness of the acquisition workforce, particularly within the Department of Defense (DoD). These orders emphasize the importance of delivering high-quality, secure software quickly through reuse, acquisition, or custom development. The key focus is on enabling the delivery of resilient software capabilities at the speed of relevance, which is crucial for maintaining a competitive advantage in modern military operations.

Another notable effort is the Software Modernization Initiative, launched under the Department of Government Efficiency (DOGE), aimed at improving the quality and efficiency of government-wide software, network infrastructure, and IT systems.

These efforts align with broader initiatives to streamline operations, reduce inefficiencies, and promote multi-discipline, multi-functional, and cross-departmental collaboration. The initiatives also emphasize the importance of modern software acquisition practices, such as the use of the Software Acquisition Pathway, which supports use of agile principles.

Defense Secretary Pete Hegseth's memo, "Directing Modern Software Acquisition to Maximize Lethality," issued March 6, 2025, states that "software is at the core of every weapon and supporting system we field to remain the strongest, most lethal fighting force in the world."

One official at DoD News, after referencing Pete Hegseth's memo, stated that "as DoD races to implement the secretary's directive, the stakes are high. … Software-defined warfare is not a future construct, but the reality we find ourselves operating in today" (Lopez & Shinego, 2025).

The most recent testimonial was presented by Sean Brady, Director, Software Acquisition & Sustainment (A&S), DoD SW Cadre Deputy Director, Joint Interoperability, March 19, 2025, at DAU's monthly "Let's Talk Agile" webinar. He reported that the newly formed DoD SW Cadre's mission is to accelerate transformation to modern SW acquisition environments. To "move out" smarter/faster, program managers recommend support from the A&S Software Enablement Team. The authors strongly recommend that program offices invest in implementing agile principles to improve technology transfer in current and future defense weapon systems.

References

Althris Training. (2024). *Agile transformation at John Deere: A case study in lean and agile practices*. https://www.linkedin.com/pulse/agile-transformation-john-deere-case-study-lean-practices-althris-fktve/

Brady, S. (2025). *Let's talk agile: Take 4 with Sean Brady software pathway policy updates*. https://www.dau.edu/events/lets-talk-agile-take-4-sean-brady-software-pathway-policy-updates

Carpenter, D. (2025, March 7). *Bolstering defense tech cybersecurity from the start*. https://www.nationaldefensemagazine.org/articles/2025/3/7/bolstering-defense-tech-cybersecurity-from-the-start

Center for Public Policy and Private Enterprise. (2012). *The Joint Tactical Radio System: Lessons learned and the way forward*. University of Maryland, School of Public Policy. https://apps.dtic.mil/sti/pdfs/ADA623331.pdf

Christ, E. (2023, October 27). Army strategizes shift to agile test and evaluation of software. https://www.eis.army.mil/newsroom/news/peo-eis-wide/army-strategizes-shift-agile-test-and-evaluation-software

Cornwell, M. (n.d.). *Defense acquisition overview: DoD defense acquisition framework (Acquisition pathways)*. https://dau365.sharepoint.com/:p:/r/sites/Class-Template-EXE4000/_layouts/15/Doc.aspx?sourcedoc=%7B538C1914-7956-42E6-BC15-2F7DF75060E8%7D&file=20241209%20CORNWELL%20AAF%20EXE4000.pptx&action=edit&mobileredirect=true

DAU. (n.d.). *FY 25 NDAA Section 832 field training teams: CONOPS review with deans and directors*. https://dau365.sharepoint.com/:p:/r/sites/ADOMA/_layouts/15/Doc.aspx?sourcedoc=%7BD8CD6C88-4E2E-43C3-8762-05AFDAFF2953%7D&file=Sec.%20832%20FTT%20CONOPS%20Overview%20for%20Deans%20&%20Directors%20(27%20Mar%202025).pptx=&action=edit&mobileredirect=true arch 27, 2025)

Defense Innovation Board. (2019). *Software is never done: Refactoring the acquisition code for competitive advantage*. https://media.defense.gov/2019/May/01/2002126690/-1/-1/0/SWAP%20EXECUTIVE%20SUMMARY.PDF

Defense Scoop. (2025, June 9–13). *Defense Tech Week website*. https://defensescoop.com/2023/04/12/us-cyber-command-requests-nearly-90m-for-offensive-platform/

Department of Defense. (2025, March 6). *Directing modern software acquisition to maximize lethality* [Memorandum]. https://media.defense.gov/2025/Mar/07/2003662943/-1/-1/1/DIRECTING-MODERN-SOFTWARE-ACQUISITION-TO-MAXIMIZE-LETHALITY.PDF

GAO. (2018). *VH-92A Program is stable and making progress while facing challenges* (GAO-18-359). https://www.gao.gov/products/gao-18-359

GAO. (2024). *DoD acquisition reform* (GAO-25-107003). https://www.gao.gov/assets/gao-25-107003.pdf

Green, C., & Hammer, C. (n.d.). *U.S. Army, Program Executive Office, Intelligence, Electronic Warfare, and Sensors website*. https://peoiews.army.mil/pm-cs/

Hillman, A. (2024, February 6). *CSC2 moves full speed ahead on the software acquisition pathway*. https://www.jpeocbrnd.osd.mil/Media/News/Article/3666146/csc2-moves-full-speed-ahead-on-the-software-acquisition-pathway/

Kowal-Swartout, P. A. (2023). *Software integration options for the F-22 and F-35 Major Defense Acquisition Programs* (Master's thesis, Naval Postgraduate School). https://hdl.handle.net/10945/72206

Lawton, D. (2023, May 19). *How agile is changing the way PEO EIS develops software*. https://www.eis.army.mil/newsroom/news/peo-eis-wide-data/how-agile-changing-way-peo-eis-develops-software

Lopez, C. T., & Shinego, W. (2025). *Modern software acquisition to speed delivery, boost warfighter lethality*. Department of Defense News. https://www.defense.gov/News/News-Stories/Article/Article/4114775/modern-software-acquisition-to-speed-delivery-boost-warfighter-lethality/

Office of the Director of National Intelligence. (2025, March). *Annual threat assessment of the U.S. intelligence community (U)*. https://www.dni.gov/files/ODNI/documents/assessments/ATA-2025-Unclassified-Report.pdf

Robinson, J., Upthegrove, M., Wassink, F., Willis, D., & Willis, T. (n.d.). *Presidential Helicopter Replacement Program (VH-92A)*. https://dau365.sharepoint.com/:p:/r/sites/Class-Template-BFM3000/_layouts/15/Doc.aspx?sourcedoc=%7B1B17FB6F-A19D-4503-8E0F-15F0F287BB2C%7D&file=VH%2092A%20Final%20Briefing.pptx&action=edit&mobileredirect=true

Rojoef, M. (April 4, 2024). *US taps consortium for Army's robotic combat vehicle software Prototype*. The Defense Post. https://thedefensepost.com/2024/04/04/us-robotic-combat-vehicle-software/

Smith, A. (2018, September–October). Agile values in the MQ-9 Reaper's software development. *Defense Acquisition Magazine*. https://www.dau.edu/library/damag/september-october2018/agile-values-mq-9-reapers-software-development

U.S. Army, Program Executive Office, Intelligence, Electronic Warfare, and Sensors. (2022, December 28). *Warfighter-centered design through agile software development to pace the threat.* https://peoiews.army.mil/2022/12/28/warfighter-centered-design-through-agile-software-development-to-pace-the-threat/

Utara, C. (2016, June 24). *Organic lead system integrator overview.* https://dau365.sharepoint.com/:p:/r/sites/CURDEV-PCO-ETM/_layouts/15/Doc.aspx?sourcedoc=%7BD39480F6-31AB-462E-A617-56CCB27756D0%7D&file=Organic%20LSI%20V002_00.pptx&action=edit&mobileredirect=true

Systematic Approach to Managing Human Trafficking Risks in Defense Contracting

MAJ Ammar M. Masoud, U.S. Army—is a Contracting Officer (51C) with 19 years of military service. He began his military career in 2005 as an Enlisted Cryptologic Linguist, later commissioned into the Signal Corps, then served as a Civil Affairs Officer with a focus on the CENTCOM Area of Responsibility (AOR). He transitioned to the Army Acquisition Corps, where he now serves as a Contracting Officer. MAJ Masoud holds a Bachelor of Science degree in Business Administration from California State Polytechnic University, Pomona, and is a Certified Federal Contracts Manager (CFCM), Project Management Professional (PMP), and Certified Contract Management Associate (CCMA). He is currently completing his Master of Science degree in Defense Contract Management at the Naval Postgraduate School, Monterey. His next assignment is with the 925th Contracting Battalion at Fort Drum, NY, supporting the Mission and Installation Contracting Command (MICC) and the 10th Mountain Division (Light Infantry).

Abstract

The Department of Defense (DoD) faces growing scrutiny over its ability to prevent human trafficking, particularly forced labor, in its overseas construction contracts. Despite promoting a zero-tolerance policy and a range of compliance measures, oversight bodies have repeatedly found that the DoD's efforts are fragmented, reactive, and insufficiently risk-informed. This study proposes the integration of the Office of Management and Budget's (OMB) Enterprise Risk Management (ERM) framework, as outlined in Circular A-123, into the DoD's Combating Trafficking in Persons (CTIP) program. Drawing on qualitative analysis of policy documents, federal regulations, oversight reports, and interviews with CTIP experts, this paper maps current CTIP practices against ERM's five core risk management phases: identification, assessment, response, monitoring, and communication. The research reveals significant gaps across the contract life cycle, particularly in pre-award planning and post-award oversight. To address these, the study presents a comprehensive ERM-integrated CTIP framework designed to shift the DoD's approach from reactive enforcement to proactive risk management. Recommendations include implementing trafficking risk screening tools, enhancing contractor vetting, standardizing monitoring practices, and improving interagency data sharing. By embedding ERM principles into CTIP processes, the proposed framework aims to better protect vulnerable laborers, strengthen contractor accountability, and ensure the DoD's contracting practices align with both ethical standards and legal mandates.

Introduction

Human trafficking—particularly forced labor within overseas defense contracts—remains a persistent challenge for the Department of Defense (DoD). The agency relies extensively on contractors for construction and support services in regions with elevated trafficking risks, where subcontractors may exploit vulnerable migrant workers through practices such as confiscating identification documents, charging excessive recruitment fees, or providing substandard living conditions (Morris et al., 2021). These violations undermine U.S. laws and international norms, disrupt mission effectiveness, and erode the DoD's strategic credibility (GAO, 2024).

Since 2002, the DoD has promoted a zero-tolerance policy, first articulated in National Security Presidential Directive 22, which mandates the prevention of trafficking, protection of victims, and accountability for perpetrators (TVPA, 2000). To implement this policy, the Combating Trafficking in Persons (CTIP) Program Management Office (PMO) was established under the Under Secretary of Defense for Personnel & Readiness. Each DoD contract includes FAR clause 52.222-50 prohibiting trafficking-related activities, and larger overseas contracts must include contractor compliance plans (GAO, 2014).

Despite these mandates, oversight bodies have identified critical shortcomings in CTIP implementation. Government Accountability Office (GAO) and DoD Insepctor General (IG) reports characterize the program as reactive and fragmented, lacking a framework to proactively identify, assess, or monitor trafficking risks across the contracting life cycle (GAO, 2021; GAO, 2024b). Compliance often takes the form of check-the-box actions—standard clauses and basic training—without structured risk analysis or enforcement mechanisms (GAO, 2024).

For example, the GAO found that contracting officials typically rely on self-certifications and standard clauses during procurement planning, with little to no risk-based evaluation. A recent review revealed that over half of applicable contracts failed to include required compliance plans (Morris et al., 2021). In the field, Contracting Officer's Representatives (CORs) frequently reported uncertainty regarding their CTIP oversight duties (GAO, 2021).

The real-world consequences of these gaps are evident. A 2019 DoD IG investigation revealed that a food services subcontractor on a U.S. base in Kuwait had subjected workers to debt bondage and illegal recruitment fees—clear CTIP violations that had gone undetected (DoD IG, 2019). Such incidents may result in contractor debarment, contract termination, reputational harm, and diplomatic fallout (Department of State [DOS], 2024; Morris et al., 2021).

The scope of this problem is magnified by the volume of DoD contracting in high-risk regions. From 2018 to 2020, the DoD awarded approximately $13.1 billion in contracts in countries classified by the State Department as Tier 2 Watch List or Tier 3—jurisdictions with weak labor protections and limited enforcement capacity (DOS, 2024; Morris et al., 2021). In these complex operational environments, traditional compliance mechanisms are insufficient. A more robust, systemic approach is needed to effectively manage trafficking risks (GAO, 2024b).

Research Purpose and Approach

To address these challenges, this study proposes integrating Enterprise Risk Management (ERM) principles, as outlined in Office of Management and Budget (OMB) Circular A-123, into the DoD's CTIP program. ERM offers a structured, organization-wide method to identify, assess, respond to, and monitor risks. Applied to CTIP, ERM can shift the DoD's posture from reactive enforcement to proactive risk management.

This research investigates the central question: How can the DoD systematically embed trafficking risk management into each phase of the contracting life cycle—from planning and solicitation through award and performance? To answer this, the study maps current CTIP practices against ERM's five core phases, identifies operational and oversight gaps, and proposes a comprehensive ERM-integrated framework to strengthen prevention, compliance, and accountability.

The methodology combines qualitative document analysis with expert interviews. Core sources include GAO and DoD IG reports, DoDI 2200.01, FAR/DFARS guidance, and OMB Circular A-123. Insights from CTIP officials and DoD contracting personnel are used to ground the findings in operational reality. The paper proceeds as follows: the Literature Review section presents a focused literature review of anti-trafficking policies and risk management frameworks. The Methodology section outlines the methodology. The Finding and Analysis section presents the research findings and the proposed ERM-integrated CTIP framework. The Conclusions and Recommendations section concludes with key recommendations and discusses the implications of adopting a risk-based approach to combat trafficking in military operations.

Literature Review

Human trafficking—defined as the use of force, fraud, or coercion to exploit individuals for labor or commercial sex—has long posed serious challenges in conflict zones and overseas contracting environments (GAO, 2024). Within the DoD, forced labor risks are most prevalent in military construction and base operations contracts, which often rely on large numbers of third-country nationals hired through complex subcontracting chains (DoD, 2023). These arrangements can obscure oversight and allow exploitative practices to persist, such as excessive recruitment fees, passport confiscation, and poor living conditions that trap workers in debt bondage (DoD IG, 2019; Morris et al., 2021).

The broader implications of labor trafficking extend beyond human rights violations. As Faruk et al. (2023) estimate, billions of dollars in illicit profits are generated annually through forced labor, particularly in conflict-adjacent regions. For the DoD, these abuses can undermine mission readiness, damage relationships with host nations, and even create local instability near U.S. bases (GAO, 2015; Hoots, 2019). In this context, labor trafficking becomes not just a humanitarian concern but a national security risk (White House, 2021).

To combat trafficking, the U.S. government has developed a comprehensive policy framework centered on the "3P Paradigm"—Prevention, Protection, and Prosecution—as established in the Trafficking Victims Protection Act (TVPA, 2000). A fourth "P," Partnership, was later added to emphasize interagency collaboration (White House, 2021). These principles shape federal anti-trafficking efforts, including those of the DoD. In operational terms, Prevention includes training and contractor vetting; Protection focuses on identifying and assisting victims; Prosecution ensures accountability for violators; and Partnership supports coordination with other agencies and NGOs (DoD, 2020a).

The DoD's CTIP Instruction 2200.01 institutionalizes this framework by mandating training, reporting mechanisms, and enforcement procedures (DoD, 2021a). Executive Order 13627 (2012), along with FAR Subpart 22.17 and DFARS 252.225-7040, further codifies anti-trafficking requirements for federal contractors, including bans on misleading recruitment, passport withholding, and charging recruitment fees (OMB, 2019).

However, the implementation of these policies has not kept pace with their intent. Oversight bodies have repeatedly identified enforcement gaps, inconsistent reporting, and limited monitoring capacity. GAO (2015) flagged insufficient controls over foreign labor use, while a 2021 report found that trafficking incidents were often underreported or poorly tracked across the command chain (GAO, 2024b). The DoD IG (2019) similarly observed that many officials lacked awareness of their monitoring responsibilities and called for improved training, guidance, and data management.

A recurring theme in the literature is the ad hoc and personality-driven nature of CTIP enforcement at the operational level. Morris et al. (2021) noted that implementation often depends on individual initiative—when a contracting officer or commander is engaged, compliance improves; when not, key safeguards may lapse. Additionally, oversight reviews have highlighted underutilization of federal tools such as the Federal Awardee Performance and Integrity Information System (FAPIIS) and the Contractor Performance Assessment Reporting System (CPARS), which could help flag contractors with prior trafficking violations (GAO, 2021). Jurisdictional barriers also complicate enforcement. Status of Forces Agreements (SOFAs) may restrict prosecution of trafficking crimes abroad, and foreign subcontractors can fall outside U.S. legal reach (Hoots, 2019), placing even more responsibility on internal oversight mechanisms.

Critically, despite strong policy mandates, there is no unified risk-based approach to managing trafficking threats in DoD contracting. Existing safeguards emphasize compliance—clauses, checklists, training—rather than proactive risk mitigation. Yet OMB Circular A-123 (2016) now requires federal agencies to implement ERM as part of their internal controls, including operational and compliance-related risks like trafficking. The Chief Financial Officers Council's ERM Playbook (2016) similarly urges agencies to identify cross-cutting risks and embed early-warning indicators into decision-making systems.

While ERM has been applied across DoD functions such as cybersecurity and major systems acquisition, it has not been extended to CTIP (GAO, 2021). As the GAO emphasized in its 2024 report, a systematic approach is urgently needed to address trafficking risks in DoD contracting environments (GAO, 2024b). Currently, the CTIP PMO lacks a risk register guidance, does not conduct recurring risk assessments trainings, and has no performance indicators beyond the required FAR dollar-value threshold that triggers compliance plan requirements (DoD, 2023).

Recent federal guidance reinforces the need for risk-based tools. OMB Memorandum M-20-01 encourages the use of trafficking risk profiles and data analytics to guide compliance oversight (OMB, 2019). The Department of Labor (DoL) has already adopted such approaches, targeting high-risk contracts based on sector and geography (DoL, 2022). International frameworks, including the UN Guiding Principles on Business and Human Rights and OECD's Due Diligence Guidance, also promote continuous risk evaluation within supply chains (OECD, 2018).

The DoD can build on internal precedents as well. Defense acquisition and systems engineering communities routinely use risk management models that incorporate structured assessment, mitigation, and monitoring cycles (Deputy Assistant Secretary of Defense for Systems Engineering, 2017). Applying these same ERM principles—risk identification, assessment, response, monitoring, and communication—to CTIP would help move the Department from reactive enforcement toward a preventative posture (OMB, 2016).

To visualize how federal agencies apply ERM principles, the GAO (2024) outlines a cyclical model that captures the essential components of risk governance. This model illustrates how agencies should align ERM efforts with mission goals, continuously identify and assess risks, determine appropriate responses, and communicate outcomes effectively. Figure 1 provides a visual representation of this ERM cycle and serves as a conceptual foundation for integrating trafficking risk into the DoD's CTIP processes.

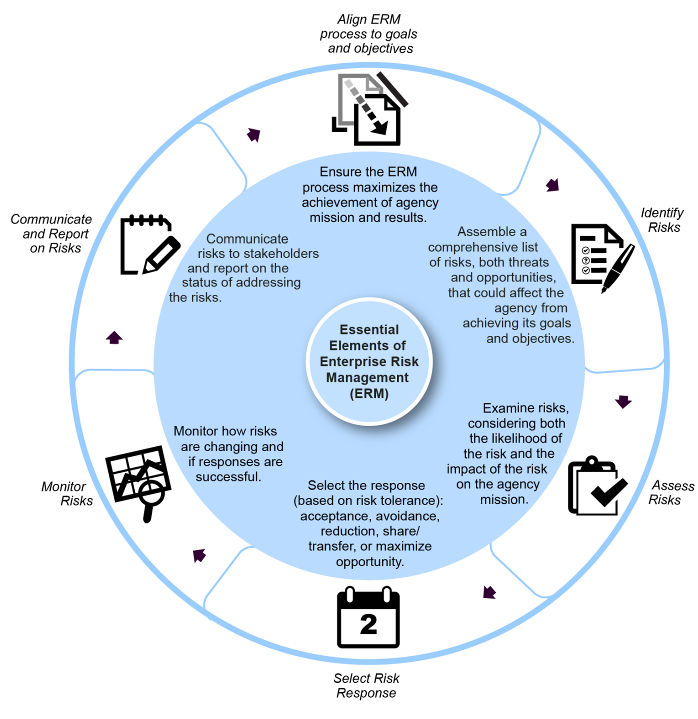


Figure 1. Essential Elements of Federal Government Enterprise Risk Management (GAO, 2024)

The literature highlights a persistent gap: the DoD has robust anti-trafficking policies but lacks a formal, enterprise-level framework to implement them consistently. The shift from compliance-oriented enforcement to a proactive, risk-informed model is essential to meet both legal mandates and operational demands (GAO, 2024; Morris et al., 2021). Integrating ERM into the DoD CTIP program would provide the structure needed to anticipate, prevent, and mitigate trafficking risks across the defense contracting life cycle.

Methodology

This study employs a qualitative, exploratory research design to assess the DoD's current CTIP practices and implementation and explore how the DoD CTIP PMO can integrate ERM principles into the CTIP program to develop a risk-informed framework aligned with these principles. The research focuses on mapping current CTIP implementation efforts across the contract life cycle—pre-award, award, and post-award—against the five phases of the ERM framework: risk identification, assessment, response, monitoring, and communication. The research methodology consisted of three components: (1) document analysis, (2) expert interviews, and (3) synthesis into a proposed ERM-integrated CTIP framework. Each component is described below.

Document Analysis

A comprehensive document review was conducted to establish a baseline of the DoD's CTIP practices, identify implementation gaps, and extract risk management insights relevant to the proposed framework. The sources analyzed fell into five major categories:

- **DoD Policies and Guidance:** Core documents included DoDI 2200.01 (2020), the DoD CTIP Strategic Plan (2014–2018), internal CTIP training materials, onboarding guides, and task force charters. These sources clarified formal roles, responsibilities, and mandated CTIP processes.
- **Acquisition Regulations:** Federal Acquisition Regulation (FAR) Subpart 22.17 and the Defense FAR Supplement (DFARS 252.225-7040) were reviewed to capture contractual requirements, including compliance plan thresholds and definitions of prohibited practices. Policy memoranda from the Defense Procurement and Acquisition Policy (DPAP) office were also included.
- **Oversight Reports:** GAO reports (2012, 2015, 2021, and 2024), DoD IG findings (e.g., DODIG-2019-088), and U.S. Central Command assessments were content-analyzed to surface recurring problems such as weak field monitoring, inconsistent reporting, and limited data tracking. These evaluations provided critical input into systemic gaps and prior recommendations.
- **Academic and Practitioner Research:** Studies by Hoots (2019), Grush (2016), Brown (2019), and Morris et al. (2021) were reviewed to incorporate external critiques and alternative conceptual frameworks addressing trafficking in military or government contracting contexts.
- **Risk Management Frameworks:** Federal guidance including OMB Circular A-123 (2016), the CFO Council ERM Playbook (2016), and OMB M-20-01 (2019) were analyzed to identify standard risk definitions, institutional best practices, and ERM application in analogous domains. The DoL's ERM Framework 3.0 was reviewed as a model for operationalizing ERM in compliance-sensitive environments.

Each document was systematically reviewed for references to ERM's five phases—risk identification, assessment, response, monitoring, and communication—and cross-coded by contract life cycle stage (pre-award, award, post-award). This dual-coding process enabled structured mapping of CTIP activities against ERM principles, highlighting phase-specific and life cycle–specific gaps. For instance, no policy documents mentioned trafficking risk assessments during acquisition planning, confirming a systemic pre-award blind spot.

Expert Interviews

To supplement the document analysis with operational insights, semi-structured interviews were conducted with subject matter experts, including DoD contracting officers, CTIP PMO staff, Defense Acquisition University (DAU) contingency contracting experts, and policy/compliance specialists. Interviews lasted approximately one hour each and were guided by a flexible questionnaire focusing on current CTIP processes, enforcement challenges, oversight practices, operational implementation gaps, and opportunities for risk-based integration.

Interviewees were selected for their subject-matter expertise and operational experience with CTIP in various geographic and functional contexts in coordination and collaboration with the DoD CTIP PMO Director. Each interview explored the following themes for a comprehensive understanding of both institutional practices and frontline challenges:

- How CTIP responsibilities are integrated (or neglected) in different phases of contracting
- Challenges in enforcing CTIP clauses and vetting contractors at the operational and tactical levels
- Suggestions for strengthening oversight or implementing risk-based practices
- Reactions to draft components of the proposed ERM-integrated framework
- CTIP experts recommendations to enhance the CTIP PMO contributions to the fight against forced labor in defense contracting

Open-ended questions encouraged real-world examples. One contracting officer observed, "Our process only kicks in after something goes wrong," highlighting the reactive nature of existing CTIP enforcement in a deployed contingency environment. Interviewees were assured anonymity and confidentiality to promote honest reflection on internal shortcomings.

All interviews were analyzed using qualitative content analysis to extract themes, corroborate gaps identified in documents, and surface practical solutions. For example, the GAO's finding that CORs were often unaware of CTIP duties was confirmed through interviews citing inadequate training and the lack of role clarity. However, a DAU professor interviewed asserted that new CORs' guidelines and trainings were updated to reflect CTIP measures on ground; nevertheless, further training and awareness are still required for CTIP implementation effectiveness by the CORs.

Data triangulation across policy documents and practitioner insights was used to validate findings and enhance reliability. The study adopts a case-study approach to contextualize risks and practices within real-world contracting environments, particularly in Tier 2 and Tier 3 countries as classified by the State Department's Trafficking in Persons Report. The integration of document analysis and expert interviews allows for a multidimensional view of current practices and the feasibility of embedding ERM into CTIP program management.

Framework Synthesis and Validation

Findings from the document review and expert interviews were synthesized to develop a proposed ERM-integrated CTIP framework. A gap analysis matrix was constructed, aligning ERM's five phases with the three key contract life cycle stages. Each cell was coded to indicate the presence (✓) or absence (✕) of relevant CTIP practices, producing a structured visualization of systemic gaps (see Table 1 in Findings and Analysis).

The framework design was refined iteratively. Suggestions from interviewees and NPS advisors and insights from ERM best practices were incorporated to propose targeted improvements—such as risk screening tools in pre-award planning and CTIP monitoring plans in Quality Assurance Surveillance Plans (QASPs). Draft framework elements were informally validated with two experts (a CTIP PMO official and NPS Advisors), who confirmed feasibility and recommended minor clarifications.

Ethical Considerations and Limitations

All sources used were either publicly available or provided with consent. Interviewees participated voluntarily, gave informed consent, and retained the right to withdraw. No personally identifiable data or victim-specific information was collected, minimizing risk to human subjects. Confidentiality was maintained in all reporting.

The study's limitations reflect its qualitative scope. Findings depend on participant perspectives and document availability. It is possible that emerging CTIP initiatives within the DoD were not captured. While the proposed framework is grounded in best practices, it remains conceptual and untested in operational settings. These limitations are acknowledged, but they do not undermine the core insight: that a structured, risk-based approach to CTIP practices and implementation is both necessary and currently lacking in DoD contracting in overseas contingency operations and forward base support.

Findings and Analysis

This section presents the study's key findings, drawn from document review and practitioner interviews, and analyzes how the DoD's current CTIP practices align with ERM principles. It begins by diagnosing the limitations of the existing CTIP process and then maps critical gaps across the contract life cycle. These findings serve as the foundation for the ERM-integrated framework proposed in the following section.

Current DoD CTIP Process and Limitations

The DoD existing process for addressing trafficking in persons (TIP) in its contracts is largely structured around reactive compliance-centered enforcement. This process delineates how TIP allegations involving defense contracts are reported and resolved, distinguishing among incidents involving Service members, DoD civilians, or contractors. When a TIP allegation arises—such as forced labor or recruitment abuse—it is typically reported through channels such as an IG hotline or the chain of command. Upon notification, the matter is referred to investigative authorities (e.g., the DoD IG or Defense Criminal Investigative Organizations), and the contracting officer is notified in accordance with FAR 52.222-50. From there, a series of accountability actions may unfold, including contract remedies (e.g., stop work orders or terminations), referrals to Suspension and Debarment Officials (SDOs), and, when applicable, criminal prosecution. This system reinforces the Department's zero-tolerance policy by imposing punitive measures when violations are substantiated.

This process is visually summarized in the Department's CTIP Case Process Flow (Figure 2), which outlines the referral and adjudication pathways for TIP incidents involving contractors, civilians, or service members, including administrative, criminal, and contractual remedies.

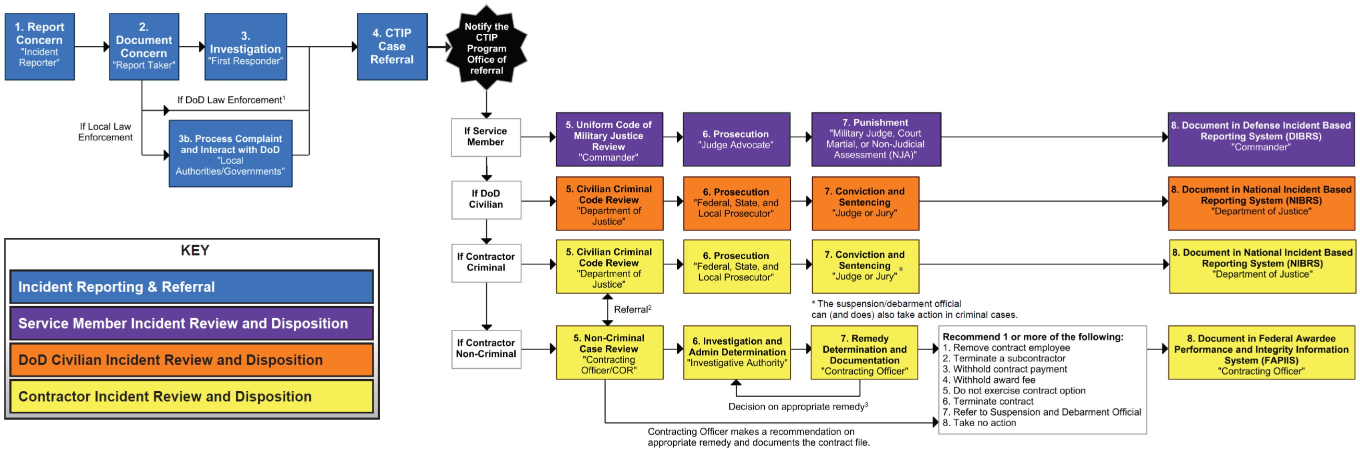


Figure 2: DoD CTIP Case Process Flow

While this process provides a clear mechanism for response, it suffers from critical limitations when evaluated through a risk management lens. Most notably, the DoD's current CTIP enforcement posture is fundamentally reactive. As one CTIP expert noted, "our process kicks in only after a problem is evident, not necessarily to prevent one" (Interview, 2024). This observation aligns with GAO findings that the Department focuses on post-violation enforcement, rather than proactively identifying and mitigating TIP risks throughout the contract life cycle (GAO, 2024).

Document analysis and practitioner interviews identified several specific shortcomings:

- **Undefined Preventive Roles and Oversight Responsibilities:** CORs are not consistently trained or tasked with CTIP monitoring. GAO (2021) reported that many CORs lacked awareness of their TIP oversight duties, a finding confirmed in interviews. A DAU professor acknowledged that CORs "lacked specific guidance on what to check or monitor regarding trafficking compliance" (Interview, 2024). The absence of explicit responsibilities and tools contributes to inconsistent field-level oversight that undermines prevention.
- **Fragmented Reporting and Data Capture:** Although multiple reporting mechanisms exist (e.g., IG hotline, chain of command, local law enforcement), they are not centrally coordinated. A 2021 GAO review noted that TIP incidents were often incompletely reported or not captured in a unified system, limiting visibility for senior leadership and precluding trend analysis. Moreover, prior contractor violations are not systematically shared across Components or contracting offices, meaning officials may award new contracts without knowledge of past TIP infractions (GAO, 2021).
- **Limited Integration of Risk Management:** The current CTIP process lacks formal mechanisms for trafficking risk identification, assessment, or prioritization at any stage of the contract life cycle. There is no structured protocol to flag high-risk contracts based on geography, sector, labor demographics, or subcontracting complexity. As one contracting officer explained, "we don't have a tool that says 'this contract has a high risk of labor trafficking' apart from the dollar threshold," highlighting the absence of early-stage risk profiling (Interview, 2024).

- **Lack of Strategic Visibility at the CTIP PMO Level:** Although the CTIP Case Process Flow instructs DoD Components to notify the CTIP PMO of trafficking referrals, this step is inconsistently executed in practice. One CTIP expert noted that "Components fail to notify the PMO when contractor-related TIP incidents occur, even when substantiated," resulting in the PMO being unaware of field-level incidents that have significant strategic implications (Interview, 2025). This breakdown hampers centralized tracking, inhibits strategic oversight, and reduces the Department's ability to identify patterns, direct resources, or report comprehensively on TIP risk trends.

Collectively, these findings reveal that while the DoD's CTIP system is capable of enforcing penalties after violations, it does not operate as a preventive risk management function. The existing framework lacks integration of TIP risk considerations into contract planning, source selection, and performance monitoring—functions that should align with the principles of ERM under OMB Circular A-123. Without a proactive, systemic approach, the Department remains vulnerable to preventable harm, missed contractors and subcontractors' patterns, and reputational damage.

Analysis of the ERM Phases in the Contract Life Cycle

Drawing on GAO reports, DoD policy documents, and practitioner interviews, this analysis examined the DoD's CTIP implementation through the lens of the ERM cycle (per OMB Circular A-123) to identify how a more proactive, risk-informed approach can strengthen the program. ERM provides an ideal analytical lens for examining CTIP implementation because it offers a systematic approach to identifying, assessing, and managing risks across an organization—precisely what the GAO has identified as lacking in current anti-trafficking efforts. Using the five ERM phases—identification, assessment, response, monitoring, and communication—each stage of the contracting life cycle (pre-award, award, and post-award) was examined for current strengths and gaps.

Risk Identification

The DoD includes standard anti-trafficking clauses (FAR 52.222-50) in relevant contracts and requires compliance plans for overseas contracts over $500,000. However, contracting officials lack tools to assess TIP risk systematically during planning. No formal risk profiling—based on country risk tiers, industry sectors, or contractor history—is conducted. Systems like FAPIIS and CPARS, which track past contractor misconduct, are underutilized for TIP-specific red flags (GAO, 2021; OMB, 2020). Interviewees confirmed that trafficking concerns are rarely discussed in pre-award meetings unless tied to prior incidents. This absence of structured identification limits the ability to prioritize resources and tailor oversight to high-risk contracts.

Risk Assessment

The DoD does not currently assign trafficking risk levels (e.g., low/medium/high) to contracts or regions. Anti-trafficking requirements are implemented uniformly, regardless of contract risk profiles. Offeror compliance plans are often reviewed for presence, not quality, and performance history regarding TIP is seldom evaluated beyond existing suspensions or debarments. Interviews revealed that source selection boards treat CTIP considerations as pass/fail items rather than factors in best-value assessments. Moreover, post-award assessments are informal and reactive, triggered only by complaints or investigations (GAO, 2024b).

Risk Response

DoD has a robust policy toolkit on paper, including contract remedies, debarment procedures, and compliance plan requirements (DoD, 2019; FAR Subpart 22.17). However, enforcement is inconsistently applied. While some commands actively use post-award conferences to reinforce CTIP responsibilities and distribute awareness materials (CTIP PMO, 2021), others overlook these steps. CORs often lack the training or explicit mandates to oversee CTIP compliance, despite the existence of updated guidance and checklists in the 2022 “Contracting Officer’s Representatives Guidebook” that could facilitate such oversight. Although tools like the Army’s CTIP Job Aid and checklists exist, they are not standardized across the DoD. Consequently, violations may go unaddressed or unreported unless prompted by an external trigger or leadership attention (GAO, 2021).

Risk Monitoring

Continuous monitoring mechanisms for CTIP are weak. Most CORs do not conduct proactive checks for trafficking indicators, and routine surveillance plans rarely include CTIP-specific tasks. While annual certifications from contractors are required, the implementation of compliance plans is not actively verified. At the program level, monitoring is fragmented—TIP data is inconsistently reported across DoD components and databases. The GAO has noted lapses in incident tracking, training statistics, and data fusion, all of which hinder effective monitoring and trend analysis (GAO, 2021; DoD IG, 2019).

Risk Communication

Internal and external communication of trafficking risks remains siloed. Field-level observations often fail to escalate to leadership due to unclear reporting roles and cultural hesitations. Information on prior violations is not systematically shared across commands or incorporated into future contract planning. Communication between contracting offices, the CTIP PMO, legal advisors, and SDOs is inconsistent, leading to missed opportunities for early intervention. Externally, the DoD contributes to interagency reports (e.g., for the Trafficking in Persons Report) but lacks transparent, public-facing disclosures on CTIP enforcement outcomes (GAO, 2024; White House, 2021).

Systematic Assessment of CTIP Implementation Across the Contract Life Cycle

The above analysis of each ERM phase can be further synthesized by systematically mapping strengths and gaps across the contract life cycle stages. Table 1 provides a comprehensive visualization of this assessment, identifying where the DoD has implemented effective measures (indicated by ✓) and where significant deficiencies remain (indicated by ✕). This systematic mapping reveals patterns across both ERM phases and contract stages, highlighting structural weaknesses in the current approach to trafficking risk management.

Table 1: Mapping of ERM Phases Across DoD Contract Life Cycle Stages—Current State Assessment

ERM Phase | Contract Stage	Pre-Award (Planning & Solicitation)	Award (Contract Finalization)	Post-Award (Performance & Close-Out)
Risk Identification	✓ Standard TIP clauses included in all relevant solicitations and contracts. ✕ No formal risk profiling of contracts for TIP vulnerabilities. ✕ Inconsistent use of past performance data (FAPIIS/CPARS) to flag prior TIP issues.	✕ Limited new identification essentially assumes risks identified during pre-award carryover. ✕ No requirement to re-check for emerging risk factors at the time of award.	✕ Reliance on ad-hoc incident reporting; no continuous surveillance plan in most contracts. ✕ Lack of field assessments to proactively detect trafficking indicators.
Risk Assessment	✕ No CTIP risk level assigned to acquisitions. ✕ CTIP plan evaluation is pass/fail, not a graded risk factor in source selection.	✕ Contractor selection does not explicitly weigh TIP risk aside from basic responsibility checks. ✓ Contractors with known severe violations are likely excluded via suspension/debarment.	✕ No routine reassessment of trafficking risk as the contract evolves. ✕ Incomplete data on incidents leads to an underestimation of risk.
Risk Response (Mitigation)	✓ Anti-trafficking clause and compliance plan requirements embedded in contracts. ✕ Quality control of compliance plans is weak. ✕ No enhanced requirements for high-risk contracts beyond the standard clause (one-size-fits-all).	✓ Post-award conferences sometimes reinforce CTIP requirements and distribute awareness materials (if done). ✓ Assignment of trained CORs and inclusion of CTIP in their duties. ✕ Not consistently executed—CTIP is often not emphasized during contract kickoff.	✓ Strong enforcement tools are available: contract termination, withhold payments, personnel removal, and S&D referrals. ✓ Some use of remedies has occurred. ✕ Monitoring-based mitigation is weak—if issues aren't detected, responses can't activate. ✕ S&D seldom used as a deterrent (few referrals), potentially limiting accountability.
Risk Monitoring	✕ No explicit CTIP monitoring plan in acquisition strategy; relies on later oversight.	✕ Minimize the time during the award, aside from ensuring documents are in order.	✓Some contracts/commands implement TIP checks as part of QA surveillance. ✕ COR oversight of CTIP compliance is often lacking due to unclear guidance. ✕ Program-level monitoring is incomplete: not all incidents are tracked in databases, and training metrics are not fully reported.
Risk Communication	✓ FAR clause and RFPs communicate expectations to bidders.	✓ Award documents and kick-off meetings (if utilized), communicate	✓ Established reporting channels (IG hotline, chain of command) for incidents.

	✕ No mechanism to communicate assessed risk level to approving officials—leadership not specifically briefed on TIP risk for acquisitions.	roles. ✕ Informal communication depends on individual CO emphasis; no standard CTIP briefing requirement exists.	✕ Gaps in upward communication: Some violations are not reported or entered into systems. ✕ Lateral communication gaps: S&D officials and others are unaware of ongoing cases. ✕ Limited external transparency on CTIP issues in contracts.

Synthesis: Need for an ERM-Integrated Framework

These findings culminate in a diagnostic table (presented in the full document) that maps current CTIP implementation across ERM phases and contract stages. It reveals strengths in enforcement tools and baseline compliance measures but highlights significant gaps in proactive risk detection, differentiated oversight, continuous surveillance, and organizational learning. As OMB Circular A-123 and Memorandum M-20-01 emphasize, enterprise risks—including those involving labor trafficking—require structured, life cycle–based management. The data strongly support the need for an ERM-integrated CTIP framework that embeds trafficking risk management throughout the contracting process. The following section presents a proposed ERM-integrated framework that addresses these gaps by embedding proactive risk management practices throughout the contract life cycle, from planning and solicitation through performance and closeout.

Proposed ERM-Integrated CTIP Framework

Building on the diagnostic findings outlined in the preceding analysis, this section proposes a comprehensive framework to embed ERM principles into the DoD's CTIP program. The proposed approach transforms CTIP from a reactive, compliance-centered posture to a proactive, risk-informed system that anticipates, mitigates, and responds to trafficking risks across the contract life cycle.

This framework aligns with the five core phases of the ERM cycle—risk identification, assessment, response, monitoring, and communication—and maps these into each phase of the contracting process: pre-award, award, and post-award. In doing so, it institutionalizes CTIP risk management as an embedded part of procurement governance, consistent with the intent of OMB Circular A-123 and OMB Memorandum M-20-01.

Pre-Award Risk Screening and Profiling

Prior to solicitation, all contracts in identified high-risk categories (e.g., those involving manual labor or performance in high TIP Tier countries) should undergo a formal Trafficking Risk Assessment. This process would be supported by a standardized tool or policy guidance developed by the CTIP PMO experts in collaboration with acquisition professionals such as the Defense Pricing, Contracting, and Acquisition Policy/Contract Policy (DPCAP/CP) office Army Contracting Command and Contracting Support Brigades (CSBs) in different Combatant Commands, such as CENTCOM and the 408th CSB. The tool, policy or guidance would draw on:

- State Department TIP Tier rankings
- Labor Department ERM best practices
- Sector-specific vulnerabilities (e.g., construction, base services, and later food services)

- Anticipated workforce composition (e.g., use of third-country nationals, local expats, etc.)
- Contractor past performance (using data from FPDS, FAPIIS, and CPARS)
- Enhanced vetting of contractors with previous trafficking violations
- Contract value, duration, and geographic complexity

Based on these inputs, contracts would be assigned a risk level (Low, Medium, High), and this designation would be documented in acquisition planning documents and a centralized CTIP Risk Register. Risk levels would directly influence solicitation design, oversight planning, and contractor expectations. This pre-award shift from uniform compliance to tailored risk stratification represents a critical advancement over the current checklist model.

Enhanced Risk-Based Solicitation and Source Selection

Building on the risk rating, the framework calls for differentiated CTIP safeguards in solicitation and award processes. For High-risk contracts, RFPs should require enhanced compliance plans that include:

- Third-party audits of recruitment agencies
- Verification of wage and housing conditions
- On-site compliance officers
- Disclosure of supply chain actors (e.g., labor brokers or subcontractors)
- Participation in certified ethical recruitment programs

Additionally, evaluation criteria should assign scoring weight to the quality of anti-trafficking measures and past CTIP performance, moving beyond the binary “compliant/non-compliant” approach. The source selection process would thus incorporate CTIP as a discriminating factor, rewarding contractors who demonstrate substantive commitment to worker protection. In extreme cases, the government may consider risk avoidance strategies, such as in-sourcing certain services or disaggregating large contracts to reduce oversight burden.

Enterprise Risk Register and Strategic Governance

At the enterprise level, the CTIP PMO would maintain a DoD-wide CTIP Risk Register—a dynamic repository that tracks risk levels, incidents, and compliance trends across contracts. All medium and high-risk contracts would be logged in this dashboard, along with real-time updates on violations, investigations, and remedial actions.

This register would be reviewed quarterly by a CTIP Risk Governance Council, comprising members from the CTIP PMO, IG offices, and senior contracting and legal officials. This forum would enable enterprise-level oversight, trend identification, and strategic resource allocation—facilitating cross-Component learning and prevention.

Role Definition and Specialized Training

To operationalize oversight at the contract level, the framework emphasizes role clarity and capacity-building. COR appointment letters for high-risk contracts would explicitly assign and enforce CTIP monitoring responsibilities as outlined in the 2022 “Contracting Officer’s Representatives Guidebook” Appendix D.7 (CTIP Checklist). In parallel, a tiered training curriculum would be introduced in collaboration with DAU:

- Basic CTIP awareness for all contracting personnel
- Advanced modules for officials managing Medium/High-risk contracts

- Scenario-based training on identifying trafficking indicators, recordkeeping, and escalation procedures

The creation of CTIP liaisons at major installations with frequent high-risk contracts would further institutionalize these responsibilities, bridging gaps between the operational and strategic levels.

Continuous Monitoring Plans

For each Medium or High-risk contract, the contracting team would prepare a CTIP Monitoring Plan as part of the Quality Assurance Surveillance Plan (QASP). These plans would include:

- Scheduled site visits (including unannounced inspections)
- Confidential worker interviews in native languages
- Verification of legal recruitment conditions (e.g., no fees, possession of passports)
- Monitoring indicators and defined escalation protocols

These activities would be formally logged, and key metrics—such as the number of worker interviews conducted or non-compliance triggers—would be reported monthly to the CTIP PMO for tracking and metrics compilation. This approach transitions CTIP oversight from ad-hoc responsiveness to deliberate, data-informed vigilance to support the CSBs in countries.

Standardized Incident Response Protocol

In the event of a suspected violation, the framework mandates a coordinated incident response workflow led by a multi-agency CTIP Incident Team. This team—comprised of contracting, investigative, legal, and victim support personnel—would ensure synchronized enforcement and reporting. Responsibilities include:

- Immediate evaluation of contractual remedies (e.g., stop-work order)
- Notification of the SDO
- Activation of victim support protocols (housing, repatriation)
- Logging actions and outcomes in the CTIP Risk Register

By introducing a case manager to oversee the incident life cycle, the process ensures timely updates to leadership and helps close the gap between field-level detection and enterprise learning.

Metrics and Performance Tracking

To institutionalize accountability, the framework introduces Key Performance Indicators aligned with CTIP risk management outcomes. Sample metrics include:

- % of High-risk contracts with CTIP monitoring plans
- Ratio of incidents detected proactively vs. reported externally
- Average response time from incident notification to remedy
- Completion rates for CTIP training among CORs on high-risk contracts

Quarterly dashboard reviews would be led by the CTIP Task Force, with results informing program adjustments and oversight briefings to leadership. This datacentric approach supports both internal accountability and external transparency.

Interagency and Industry Collaboration

Recognizing that TIP risk in government contracting transcends the DoD alone, the framework promotes external engagement through:

- Active participation in the OMB-led Procurement & TIP Task Force
- Annual industry CTIP roundtables to align expectations and share best practices
- Data-sharing partnerships with agencies like USAID and State Department on common vendors and country risks

This collaborative stance positions the DoD not only as a policy implementer, but as a leader in anti-trafficking innovation across the federal landscape.

Illustrative Scenario

To visualize the ERM-CTIP implementation, let's consider a $200 million base support contract in Southeast Asia. Pre-award risk screening flags the procurement as "High risk." The RFP mandates a robust CTIP plan, and past violations influence source selection. Upon award, the COR receives CTIP-specific oversight responsibilities, and a monitoring plan includes quarterly site visits and monthly worker interviews. When concerns emerge, the Incident Team responds swiftly—triggering contract remedies, updating the risk register, and briefing senior officials. Later, trend analysis across contracts reveals patterns tied to a problematic subcontractor, prompting preemptive scrutiny on future awards.

Expected Gains from Implementation

The integration of ERM principles into the DoD's CTIP program is not merely a procedural shift—it represents a fundamental transformation in how trafficking risks are understood, addressed, and governed. This section outlines the practical benefits that implementation of the proposed framework is expected to yield. Drawing from both the diagnostic findings and best practices in federal risk management, these anticipated gains span from early prevention and stronger oversight to improved coordination and long-term institutional resilience.

- **Proactive Prevention:** Upfront risk profiling and enhanced requirements stop many violations before they occur, shifting the program from incident response to risk anticipation.
- **Visibility and Accountability:** CTIP performance becomes trackable, auditable, and visible to leadership—closing gaps in oversight and strengthening institutional responsibility.
- **Improved Detection and Timely Response:** Proactive site-level monitoring, standardized incident protocols, and well-trained personnel result in earlier detection, swifter remedies, and stronger deterrents.
- **Contractor Incentives and Culture Change:** Contractors understand that the DoD prioritizes CTIP. Competitive pressures and transparent evaluation criteria encourage long-term investment in ethical labor practices.
- **Alignment with Federal Risk Governance:** By integrating TIP as a strategic risk, the framework supports OMB A-123 compliance, responds to GAO recommendations, and models interagency best practices.
- **Mission Resilience and Operational Continuity:** Preventing labor

exploitation minimizes performance disruptions and reinforces ethical and operational credibility of U.S. defense missions abroad.

Implementation Considerations

Transitioning to this model will require resourcing and cultural shift. Challenges may include increased training demands, contractor pushback, and the need to pilot tools (e.g., risk assessment checklists) before full-scale deployment. However, the strategic and ethical benefits outweigh the transitional costs. Importantly, this framework is designed to evolve—integrating seamlessly with the DoD's digital modernization. Risk registers can be embedded into existing procurement systems, and data analytics (e.g., anomaly detection via AI) can strengthen future risk identification.

Applying ERM to CTIP is not only viable—it is urgently necessary. This framework offers a structured path to embed anti-trafficking oversight at every level of the defense acquisition system. Through risk-driven prevention, enhanced contractor accountability, and interagency collaboration, the DoD can fulfill its zero-tolerance policy with integrity, foresight, and leadership.

Conclusions and Recommendations

This study set out to determine how the DoD can systematically address human trafficking risks within its contracting processes. Through a comprehensive analysis of current practices and the application of the OMB Circular A-123 ERM framework, it has become evident that integrating risk management into the DoD's CTIP program is both a critical need and a viable path forward. While the DoD has established anti-trafficking policies, significant gaps remain in execution, leaving vulnerable workers at risk and the Department exposed to unethical and unlawful practices in its supply chain. By adopting an ERM-integrated CTIP framework, the DoD can transition from a reactive stance to a proactive posture, embedding trafficking risk considerations into everyday contracting activities.

Key Recommendations for the DoD's CTIP Program

To operationalize the ERM-integrated CTIP framework, this section outlines a set of targeted, actionable recommendations aimed at institutionalizing trafficking risk management across the Department's acquisition life cycle. These recommendations are grounded in the findings of this study and aligned with federal risk management guidance. Together, they provide a roadmap for transforming CTIP from a compliance obligation into a strategic function that safeguards both mission integrity and human rights.

1. **Integrate ERM into CTIP Policy and Guidance:** The DoD should formally incorporate ERM principles into its CTIP directives and related acquisition policies. This includes updating DoD Instruction 2200.01 (CTIP) and the DFARS to mandate risk-based approaches, such as requiring risk assessments for contracts in designated high-risk categories and maintaining a CTIP risk register that feeds into ERM reporting. The CTIP PMO should develop implementation guidance for field offices on conducting risk profiling and developing monitoring plans, aligned with the OMB ERM framework.
2. **Pilot the Framework in High-Risk Environments:** Initiate a pilot program applying the ERM-integrated CTIP framework in high-risk settings, such as contracts within the U.S. Central Command (CENTCOM) Area of Responsibility (AOR) involving large overseas service contracts. This pilot would implement all elements of the framework on select contracts over a 1–2 year period, allowing for refinement based on real-world feedback and demonstrating proof of concept. Leadership support is crucial, potentially through a policy memo from the Office of the Under Secretary of Defense for

Personnel and Readiness or Acquisition and Sustainment.

3. **Strengthen Contract Requirements and Clauses:** Enhance anti-trafficking provisions in the FAR/DFARS to enable a risk-based application. Introduce clauses requiring additional safeguards for high-risk contracts, such as detailed compliance plans, independent audits, and annual CTIP compliance reporting from contractors. Mandate that prime contractors flow down CTIP monitoring obligations to subcontractors, ensuring active oversight beyond current requirements. Update source selection procedures to allow evaluation credit for superior CTIP practices, providing regulatory clarity.
4. **Implement a CTIP Risk Dashboard for Decision-Makers:** Develop a unified CTIP risk dashboard accessible to key decision-makers, displaying metrics such as the number of high-risk contracts by command, status of monitoring plans, and incidents reported and resolved. Integrate this dashboard into existing contracting data systems, like the Procurement Business Intelligence Service, or as an extension of the DoD's internal human trafficking case management system. This visualization operationalizes the risk register concept, keeping leadership attention on the issue.
5. **Enhance Training and Resources for Oversight Personnel:** Invest in specialized training for contracting officers, CORs, and contract administrators involved in medium to high-risk contracts. Training should cover identifying trafficking red flags, interviewing workers, documenting and reporting findings, and leverage case studies from past incidents. Provide practical tools, such as CTIP Monitoring checklist templates, guides for evaluating compliance plans, and a library of best practice compliance plan examples. Ensure support for oversight personnel, including access to CTIP experts and language support for worker interviews.
6. **Improve Interagency Coordination and Share Data:** Formalize information-sharing protocols with other agencies, particularly the Department of State's TIP office and the DoL. Share information on recruitment agencies involved in abusive practices to inform watchlists and vice versa. Propose regular interagency meetings focused on trafficking in federal contracting, ensuring lessons learned inform broader strategies. Engage with host nation authorities through embassies to strengthen enforcement, coordinating on inspections of labor camps.
7. **Continuously Refine the Framework with Data Analytics and Feedback:** Treat the CTIP risk management framework as a living program, evolving through data analysis and field feedback. Analyze collected data to identify effective strategies and areas needing improvement. Incorporate modern data analytics or AI to enhance early-warning capabilities, such as mining contract performance reports or using network analysis to identify problematic subcontractors. Establish channels for field personnel to suggest improvements or report obstacles and conduct formal reviews after full implementation to make necessary adjustments.

Implications for the DoD CTIP PMO

Adopting the proposed framework will transform the CTIP PMO from a policy and training overseer into a dynamic risk management coordinator. This transition requires developing new capabilities in data analysis and program management, potentially staffing with personnel experienced in risk management, auditing, data analytics and compliance. The PMO would manage the CTIP risk register/dashboard and become the focal point for reporting CTIP risk status to DoD leadership. Strengthening collaboration with contracting policy offices, the DoD IG, and Combatant Command representatives is essential, possibly through establishing a dedicated ERM integration working group. By proactively managing risks, the CTIP PMO will enhance the DoD's ability to prevent issues, efficiently allocate resources, and respond effectively to incidents.

Broader Significance and Future Outlook

Implementing the ERM-integrated CTIP framework extends beyond improving contract oversight; it reinforces the DoD's commitment to human rights and mission effectiveness. By eradicating forced labor from its overseas contracts and supply chain, the DoD upholds U.S. values, protects workers supporting military missions, and strengthens the moral legitimacy of U.S. operations. This initiative can serve as a model for integrating ERM into other cross-cutting issues, such as vendor integrity, contract fraud, and human rights due diligence in global contracting. Future research should explore adapting the framework to manufacturing supply chains, leveraging supply chain transparency technologies, developing quantitative methods to measure risk reduction, comparing interagency frameworks, and conducting cost-benefit analyses to justify investments in anti-trafficking measures.

References

ACQ.gov. (2025, January 17). *52.222-50 Combating trafficking in persons*. Acquisition.gov. https://www.acquisition.gov/far/52.222-50

Brown, M. C. (2019). *21st century and no emancipation in sight: Analyzing human trafficking policy and spending in Department of Defense procurement* (Technical Report NPS-CM-19-XX). Naval Postgraduate School.

Chief Financial Officers Council. (2016). *Playbook: Enterprise risk management for the U.S. federal government*. U.S. Government Publishing Office.

Combating Trafficking in Persons Program Office. (2024). *Acquisition training*. U.S. Department of Defense. https://ctip.defense.gov/Training/Acquisition-and-Contracting-Personnel/

CTIP Program Management Office. (2024, May). *Quarterly CTIP newsletter: May 2024*. Department of Defense Combating Trafficking in Persons Program Office.

DoD. (2014). *Strategic plan for combating trafficking in persons (CTIP) 2014–2018*.

DoD. (2019). *CTIP onboarding packet for points of contact and task force members*.

DoD. (2020). *DoD Instruction 2200.01: Combating trafficking in persons (CTIP)*. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/DODi/220001p.pdf

DoD. (2021). *Combating trafficking in persons (CTIP): TIP laws, policies, and regulations*.

DoD. (2022). *Contracting officer's representatives guidebook*. Office of the Under Secretary of Defense for Acquisition and Sustainment. https://www.acq.osd.mil/asda/dpc/cp/policy/other-policy-areas.html

Department of Defense Inspector General. (2019). *Evaluation of DoD efforts to combat trafficking in persons in Kuwait* (Report No. DODIG-2019-088). https://media.defense.gov/2019/Jun/11/2002149809/-1/-1/1/DODIG-2019-088.PDF

Department of Justice. (2021). *Attorney General's annual report to Congress on U.S. government activities to combat trafficking in persons*. https://www.justice.gov/d9/2023-

04/AG%20HT%20Report%20FY21FINALPDF.pdf

Department of Labor. (2022). *Framework for enterprise risk management* (Version 3.0). Office of Inspector General. https://www.oig.dol.gov/public/reports/FINAL__12-January-2022__OIG%20Strategic%20Plan%202022%20-%202026.pdf

Department of Labor. (2024). *Step 2: Assess risks and impacts*. https://www.dol.gov/agencies/ilab/comply-chain/steps-to-a-social-compliance-system/step-2-assess-risks-and-impacts

Department of State. (2024). *2024 Trafficking in persons report: Kuwait*. https://www.state.gov/reports/2024-trafficking-in-persons-report/

Deputy Assistant Secretary of Defense for Systems Engineering. (2017). *Department of Defense risk, issue, and opportunity management guide for defense acquisition programs*. https://www.acq.osd.mil/se/pg/guidance.html

Dixon, L., Livingston, D., Long, W. E., Jr., & Yoder, E. C. (2024, August). Guarding against human trafficking in DoD contracting. *Defense Acquisition Research Journal*. https://www.dau.edu/library/damag/july-aug2024/guarding-against-human-trafficking

Faruk, A., Ibrahim, D. T., & Danjuma, I. F. (2023). Trafficking in human: A modern-day slavery against the third world states. *African Journal of Law, Political Research and Administration, 6*(2), 1–21. https://doi.org/10.52589/AJLPRA-X7K8ZAXZ

GAO. (2012). *Suspension and debarment: DoD has active referral processes, but action needed to promote transparency* (GAO-12-932). https://www.gao.gov/products/GAO-12-932

GAO. (2015). *Human trafficking: Oversight of contractors' use of foreign workers in high-risk environments* (GAO-15-102). https://www.gao.gov/products/gao-15-102

GAO. (2021). *Human trafficking: DoD should address weaknesses in oversight of contractors and reporting of investigations related to contracts* (GAO-21-546). https://www.gao.gov/products/gao-21-546

GAO. (2024a). *Federal contractors: Actions needed to improve quality of performance and integrity data* (GAO-24-106911). https://www.gao.gov/products/gao-24-106911

GAO. (2024b). *Human trafficking: Agencies need to adopt a systematic approach to manage risks in contracts* (GAO-24-106973). https://www.gao.gov/products/gao-24-106973

Grush, C. M. (2016). *Human trafficking: The United States military's hollow policies* [Master's thesis, Marine Corps University]. Defense Technical Information Center.

Hoots, A. B. (2019). Severing the connection between sex trafficking and U.S. military bases overseas. *Fordham Law Review, 88*(2), 733–768.

Johnson, C. (2019). *News from CTIP program manager in Afghanistan*. U.S. Department of Defense. https://ctip.defense.gov/portals/12/Quarterly%20CTIP%20Newsletter%20Oct%202019_vol4_1.pdf

Morris, A. L., Crouch, W. C., IV, & Peaslee, K. P. (2021). *The price of slavery: An analysis of human trafficking policy and spending in Department of Defense procurement* [MBA professional project, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/67882

Organisation for Economic Co-operation and Development. (2018). *OECD due diligence* *guidance for responsible business conduct*. https://mneguidelines.oecd.org/

Office of Management and Budget. (2019). *Anti-trafficking risk management best practices & mitigation considerations* (OMB Memorandum M-20-01). Executive Office of the President. https://www.whitehouse.gov/wp-content/uploads/2019/10/M-20-01.pdf

Office of Management and Budget. (2020). *Management's responsibility for enterprise risk management and internal control* (OMB Circular No. A-123, M-16-17). Executive Office of the President. https://www.whitehouse.gov/wp-content/uploads/legacy_drupal_files/omb/memoranda/2016/m-16-17.pdf

Office of Management and Budget. (2021). *Federal performance framework: Implementation guidance*. Executive Office of the President.

The White House. (2002). *National security presidential directive 22: Combating trafficking in persons*. Executive Office of the President.

The White House. (2021). *The national action plan to combat human trafficking*. Executive Office of the President. https://www.whitehouse.gov/wp-content/uploads/2021/12/National-Action-Plan-to-Combat-Human-Trafficking.pdf

Trafficking Victims Protection Act of 2000, Pub. L. No. 106-386, 114 Stat. 1464 (2000).

U.S. Central Command. (2019). *Evaluation of DoD contracts regarding combating trafficking in persons (CTIP)*. U.S. Department of Defense.

Enhancing Energy Resilience in Navy Region Southwest: Bridging Capability and Capacity Gaps with Modern Technologies and Utility Acquisition Approaches

Eric Hahn—Eric Hahn's 27-year military career in the United States Marine Corps and Navy involved facilities engineering, acquisition, and logistics duties while leading and developing military and civilian, technical, and acquisition support personnel. He has overseen Navy energy-related projects worldwide and was responsible for managing energy and utilities infrastructure supporting industrial activity at a Naval Surface Warfare Center. He holds engineering degrees from the University of Rhode Island, Rochester Institute of Technology, and the University of Hawaii, and three graduate certificates from the Naval Postgraduate School, and is a registered Professional Engineer.

Guy Warner—Guy Warner received his BA in government from St. Lawrence University and completed PhD-level course work in economics at George Washington University. Utilizing his decades of experience developing project financing and optimization models for environmental mitigation projects in reforestation, energy efficiency, and renewable energy, he founded Pareto Energy in 2004 in collaboration with the United States Conference of Mayors ("USCM"), where he now sits on the Board of U.S. Mayor Enterprises.

Navid Gohardani—Navid Gohardani holds PhD and Licentiate of Engineering degrees in civil and architectural engineering, a Master of Science degree in civil engineering, Bachelor and Master of Business Administration degrees, and a Bachelor of Science degree in the built environment. With over two decades of professional experience in the Architecture, Engineering, and Construction (AEC) industry, and now serving as an Interdisciplinary Engineer/Construction Manager with the Facilities Engineering and Acquisition Division (FEAD) at Naval Facilities Engineering Systems Command (NAVFAC), he is well versed in working with sustainable building, lean construction, and energy efficiency projects, particularly in the Research and Development stages.

Abstract

Naval installations and defense communities are Centers of Gravity (COGs), enabling missions across warfare enterprises. However, they're susceptible to interruptions due to vulnerable, deteriorating energy infrastructure, lack of distributed energy resources, rapidly growing demand for power, and reliance on the existing utilities' central "macro-grid" that cannot meet surging demands for high quality electric power. These difficulties underscore the need for prompt and innovative energy solutions that are secure, affordable, and acceptable. This paper explores the multifaceted challenges and opportunities faced by defense energy managers and public works professionals in understanding predominant capability and capacity gaps and the planning and acquisition of utilities from integrated defense community energy systems to address them. This study combines guidelines from Commander, Navy Installations Command, to create educational content for energy managers and a thorough framework for installation energy and utility gap identification and mitigation, incorporating modern applications of technologies such as interconnected microgrids and Digital Twins (DTs). This study offers focused solutions inspired by effective deployments of Integrated Community Energy Systems (ICES). It also introduces a four-level engineering and institutional planning and project delivery methodology. For wider use across Navy installations, the conclusions seek to ensure operational resilience and scalability at selected installations.

Introduction

The U.S. Southwest region hosts several critically important Navy and Marine Corps installations. This is obvious in San Diego, but also is true in areas outside San Diego, within Southern California Edison (SCE) electric utility service territory, as well as other areas. Key installations in SCE territory include Naval Weapons Station Seal Beach, responsible for loading, storing, and maintaining Navy weapons, and its detachments across California. Naval

Base Ventura County serves as a major aviation shore command and mobilization base for Naval construction forces, providing airfield, seaport, and base support services. Additionally, Marine Corps Air Ground Combat Center (MCAGCC) at Twentynine Palms focuses on large-scale live-fire training and operational readiness. These installations collectively support missions such as fleet logistics, weapons maintenance, aviation operations, and advanced combat training, ensuring operational readiness and strategic capabilities for the Navy and Marine Corps in the region (MilitaryBases.com, n.d.).

These installations, their critical missions, the surrounding defense communities that support them, and the utilities that are their critical enablers make up a complex socio-technical system that can be considered a COG. From a strategic military perspective, a COG is a primary source of strength, balance, or stability that enables a force to maintain its freedom of action, physical strength, or will to act. It can be physical, moral, or both, and exists at strategic, operational, and tactical levels of war. The concept, introduced by Carl von Clausewitz, emphasizes targeting the "hub of all power and movement" to disrupt an adversary's ability to achieve objectives. Analyzing COG using frameworks, such as Joseph Strange's critical capabilities, requirements, and vulnerabilities, is not the primary focus of this paper. However, it underscores the importance of recognizing COGs and how they might be exploited effectively by an enemy (Clausewitz, 2009; Joint Chiefs of Staff [JCS], 2020; Strange, 1996). Hence, it is of paramount importance for defense energy managers to understand the vulnerabilities of utility systems (particularly electric utilities) in the context of capability and capacity gaps and how to address them with timely and appropriate acquisition and procurement approaches like what is presented in this paper.

Aging and Vulnerable Energy Infrastructure

Many large central electric utilities in the United States (e.g., SCE) are characterized by aging utility "macro-grids," sprawling networks of power plants, transmission lines, and substations. They also face mounting challenges in delivering reliable power due to physical, cyber, and environmental threats. Built largely in the 1960s and 1970s, these grids are reaching or exceeding their 50- to 80-year lifespans (Smart Electric Power Alliance [SEPA], 2024). Electromechanical substations, critical nodes for voltage transformation, are estimated to be beyond their peak age, with an estimated 70% of U.S. transformers and transmission lines that are over 25 years old (SEPA, 2024; U.S. Department of Energy [DOE], 2021).

Physical threats, such as vandalism and terrorism, have surged, with 95 human-related electric disturbances reported in the first half of 2023 alone (American Society of Civil Engineers [ASCE], 2025). Cyberattacks, like ransomware or state-sponsored hacks, target substations' automation systems, risking widespread outages costing up to $1 trillion in economic damage (Lloyd's, 2015, as cited in Power Grid International, 2016). Environmental threats include wildfires and storms that damage infrastructure, with weather-related outages doubling from 2000–2009 to 2014–2023 (ASCE, 2025).

Replacement of aging substation components, such as high-voltage transformers, has lead times of at least 18–36 months due to supply chain constraints and manufacturing delays (National Academies Press, 2012). These delays exacerbate downtime risks, with U.S. utilities reporting a 20% increase in outage duration (System Average Interruption Duration Index) from 2006 to 2023, partly due to equipment failures (Lawrence Berkeley National Laboratory [LBNL], 2012; SEPA, 2024). Prolonged outages disrupt critical services, increasing public health and economic risks, with costs of reliability events estimated at billions annually (LBNL, 2012).

Peak energy demands could escalate as much as 40% by 2045 in SCE's service territory (Castaneda & Ioan, 2024). As energy demands escalate and threats to critical infrastructure persist, U.S. Navy installations in the Southwest, such as Naval Base Ventura

County (NBVC), face significant challenges in maintaining mission-critical operations. Deferred maintenance on SCE transmission lines heightens fire hazards and grid instability, exacerbating risks for installations in their service territory, while widespread Public Safety Power Shutoffs (PSPS) to mitigate wildfire risks mean widespread outages. Recent grid failures in SCE's service territory highlight these vulnerabilities. In February 2025, SCE reported at least 530,000 customers were without power for days from wildfire-related damage. The wildfire itself may be attributable to SCE equipment as an ignition source. Restoration was delayed in some areas for weeks (Southern California Edison, 2025).

At the outer edges of SCE's network in Port Hueneme, Ventura, and Oxnard, a severe storm in late December 2023 destroyed the Port of Hueneme's shoreside power substation, a critical system enabling docked vessels to use electricity from shore. Flooding from the storm inundated the substation, necessitating its replacement, which in the case of typical electromechanical substation technology requires several years. This long-term outage has forced vessels to rely on their engines, increasing emissions and disrupting compliance with California Air Resources Board (CARB) regulations.

The Port of Hueneme's critical role in regional logistics, handling well over $10 billion worth of goods annually, and its proximity to NBVC amplify the operational risks of such grid failures. However, critical issues prevent widespread adoption of Distributed Energy Resources (DER) that are needed now to provide more reliable, resilient, high quality power, while also deferring expensive and long-lead-time central utility distribution system upgrades. Backup power systems alone are not a profitable venture for attracting third party financing, and they are typically unable to seamlessly sustain critical loads without interruption, threatening mission readiness. Poor shore power quality, exacerbated by events like the 2023 substation destruction, damages ship equipment and forces reliance on shipboard engines, contributing to port emissions. These challenges underscore the need for innovative energy projects that can provide secure energy solutions, accepted by the surrounding community, and aligned with the Navy's Shore Energy Program. This starts at a fundamental level with addressing the critical issues preventing widespread DER adoption.

Gaps Due to Critical Issues With DER Adoption

Widespread DER adoption could defer the immediate need for expensive and slow-to-deliver central utility distribution system upgrades. However, there are three critical issues that prevent the widespread adoption of DER. These are:

1. Hosting capacity limitations
2. The difficulty of solving centralized control and optimization algorithms and
3. The reluctance of customers to adopt large-scale distributed energy resources even where the technical and economic designs seem favorable to them.

Hosting more DER can supply the increased peak demand growth SCE predicts, beyond the capacity limits of the current distribution network. This capability is known as a "dynamic hosting capacity" (Castaneda & Ioan, 2024). Such capability can help utilities manage the forecasted growth in electricity demand for things like data centers, electrified ports, and increasing shipyard production capacity. A DT model of a substation is needed to simulate real-time environments and validate control architecture use-cases to ensure interoperability between all control architecture systems prior to field deployment using actual customer DER.

It is critical that a Distribution System Operator (DSO) has the capability for control and optimization in the dispatch of large numbers of customer inverters independently. However, simulations of large numbers (~10,000) of DER have demonstrated that it is not possible to individually optimize them all using a centralized control approach. Furthermore, despite the coming peak demand increases, critical infrastructure operators are reluctant to invest in DER. This is likely attributable to inconsistencies between what existing central utilities say is needed to meet the demand in terms of DER adoption versus what they will actually allow to interconnect. Therefore, SCE (and other central utilities), along with the installations they serve, face both a capability and capacity gap with respect to hosting a large amount of DER.

Because of these persistent DER adoption issues, critical defense infrastructure operators like installation public works officers and the energy managers that advise them continue to remain dependent on external central utilities. It also can cause them to self-limit their thinking of the solution space for resilience to only things like demand management (e.g., efficiency projects to reduce their reliance on the utility macro-grid) and backup power (e.g., microgrids for "islanding" in the event of utility macro-grid failure). They are less aware of how modern Integrated Community Energy Systems (ICES) microgrids operated internally or locally with distributed utility business approaches address the critical issues of DER adoption and can meet significant expected demand growth reliably, affordably, and acceptably.

Understanding ICES

ICES is not a new concept. It broadly refers to localized energy systems integrating DER to meet community needs (electricity, heat, cooling, mobility) with goals of efficiency, sustainability, and resilience. The ICES concept has evolved over decades, shaped by technological, policy, and social shifts, with microgrids playing a significant role in modern iterations. Post–World War II energy scarcity and Denmark's cooperative tradition originally spurred these kinds of decentralized energy solutions. Early district heating systems (1940s–1950s) used waste heat from small coal/oil plants (1–5 MW) to serve towns, laying ICES groundwork. In response to the oil crisis in the 1970s, the DOE's Community Systems Program aimed to reduce U.S. energy imports and sponsored a 1977 financial overview of ICES. This study formalized ICES as integrated DER systems (solar, combined heat and power [CHP], storage) for self-sufficiency. There were several U.S. examples of ICES projects using distributed CHP plants to serve areas like Starrett City in New York City which includes residential towers, commercial spaces, and community facilities. European implementations of ICES from the 1980s–2010s in rural applications (e.g., Wildpoldsried and co-ops) integrated renewables and modern technologies including some power electronics. Modern urban ICES implementation (2010–2015) leverage microgrid technology and medium voltage power electronics interconnection. When these are developed as primary power projects and operated by regulated distributed utilities, they can offer higher rates of return than microgrid projects developed for energy savings or power backup only. The evolved distributed energy business models developed at 36 European ICES have been detailed in Reis et al. (2021). Similarly, business models for numerous ICES energy cooperatives have been detailed in Kubli (2023).

Beginning in early 1992, electric power experts at Pacific Gas Electric Company (PG&E) began collaborating with their counterparts at the DOE's National Renewable Energy Laboratory (NREL), the Electric Power Research Institute (EPRI), and the Pacific Northwest National Laboratory (PNNL) on the Distributed Utility Valuation Project (Pupp,1993). The research about the mutual benefits of the distributed utility business model culminated in 1997 with a special issue of *The Energy Journal*. From numerous articles therein about applied technologies, institutional innovations and regulatory reforms, it became clear that distributed energy resources could compete with the utility macro-grid by offering less expensive power with higher levels of quality, reliability and environmental sustainability (Smeers et al., 1997).

The 1992 Distributed Utility Valuation Project demonstrated the value of "non-synchronous" interconnection using AC to DC to AC power conversion at a PG&E distributed generation system. The conclusion by 1993 was that

> Modern solid-state inverters have very reliable, fast, and sophisticated protective devices built in, are digitally controlled and hence flexibly programmable, and produce very clean waveforms. This experience and others like it confirm that "most interfacing issues are resolved or resolvable with state-of-the-art hardware and design," …and that the literature "does not reveal any unsolvable technical problems," so "In the near-term, it appears that there are no technical constraints that impede the integration of intermittent renewable technologies into...utility systems."…Distributed generators that feed the grid through appropriately designed DC-to-AC inverters can provide the desired real-time mixture of real and reactive power to maximize value. (Lovins, 2002, quoting from Wan, 1993)

PNNL researchers on the 1993 PG&E Distributed Utility Valuation Project had concluded by 1996 that the self-commutating inverter was the "technology of choice" for interconnecting large amounts of storage to the utility grid (Donnelly et al., 1996).

In the face of increasing congestion on the wide-area transmission and distribution network and surging demand for power to support artificial intelligence and the electrification of transportation, heating and cooling, SCE has accepted the need for DER as a faster, more affordable alternative to expanding the wide-area SCE transmission and distribution network (Castaneda, 2022). In 2022, SCE advised that

> The market size for DERs have grown exponentially over recent years and is projected to continue growing due to benefit awareness and realization of societal benefits, evolving rates & tariffs, decreased cost of equipment, and government incentives & policies" and that "In the future, DER aggregators … will be incentivized to compete with bulk-system energy supplied through substations and play a critical role in managing most of the DER customers on the network. (Castaneda, 2022)

Later, SCE also confirmed through testing the superior affordability and reliability of "non-synchronous" (DC link) power electronics interconnection and control (SCE, 2023).

It's important to note that "non-synchronous" (DC link) interconnection can result in dramatic cost savings. A federally funded <u>synchronous</u> control system for the 14 MW of combined heat and power at Naval Shipyard Portsmouth cost <u>$4.14 per watt</u> (Environmental Security Technology Certification Program [ESTCP], 2017). <u>"Non-synchronous" (DC link) power electronics</u> control systems are being offered at <u>$1.20 per watt</u> and, if funded by a third party, could earn an investment tax credit that results in a net price of $0.72 per KW. Moreover, the non-synchronous interconnection can earn more than the synchronous interconnection in sales of voltage and frequency regulation to support the utility-owned macro-grid. The $0.72 investment can usually be recouped in less than two years from such revenue (Pareto Energy, 2023).

Three implementations of federally funded microgrids at Naval Shipyard Portsmouth, Naval Shipyard Norfolk and the Naval Research Lab in the District of Columbia cost between $5.50 and $11.10 per watt. A third party funded microgrid to serve a hospital and federal offices in the District of Columbia cost $0.66 per watt after federal tax credits and a grant, a price that reduced cost of power by 40% as compared to power from the utility grid (Pareto Energy, 2023).

In conclusion, SCE's acknowledgement of the critical need for more DER customer adoption and acceptance of a power electronics control platform make the organization of ICES to serve installations and the community of critical infrastructure and labor that support installations a timely and affordable option for energy resilience and reliability. However, before energy managers can understand how modern ICES address DER adoption issues, they must understand ICES and key related technologies and theory. There's key terminology that's particularly relevant to understanding modern ICES.

Multi Agent Systems (MAS): MAS consist of multiple decision-making agents which interact in a shared environment to achieve common or conflicting goals. An ICES is a MAS where the energy producing or using devices and energy distribution networks are the agents and interconnected with one another. To optimize its operations, an ICES must provide near instantaneous balancing of voltage and frequency as consumer demand for power continuously changes. It is important to note here that SCE has admitted that they cannot centrally control multiple agents fast enough to respond to demand without a $400 million investment in computing technology. However, Naval Postgraduate School (NPS) and their Cooperative Research and Development Agreement (CRADA) partner, Pareto Energy, have recommended that SCE implement a more affordable and effective distributed control and optimization framework that has been developed and tested on numerous ICES. For a summary of the advantages of distributed versus centralized ICES control, see Cheng et al. (2018).

Common Pool Resource (CPR): A CPR is a framework in which MAS decision-making is made more by active consumers that use the MAS and the labor that builds the MAS, and less by governments or markets. The value of organizing an ICES as a CPR is that it enables decentralized optimization algorithms fast enough to balance the ICES. Currently, centralized optimization by a utility company cannot optimize an ICES fast enough, as noted in the DER adoption issues. Fortunately, Nobel Prize winning economist Elinor Ostrom developed a practical method for designing a CPR that she applied successfully to numerous MAS in the field. It is known as the Institutional Analysis and Development framework (IAD). Professor Ostrom (2005, 2010) summarized the benefits of consumer collective self-governance of a CPR:

> Public policies based on the notion that all CPR consumers are helpless and must have rules imposed on them by either markets or the government can destroy institutional capital that has been accumulated during years of experience in particular locations. An in-depth analysis of their experience can deepen one's appreciation of human artisanship in shaping and reshaping the very situations within which individuals must make decisions and bear the consequences of CPR use on a day-to-day basis. Success in starting small-scale initial institutions enables a group of individuals to build on the social capital thus created to solve larger problems.

Digital Twin (DT): An ICES DT is a virtual representation of a MAS that spans its life cycle, it's updated from real-time data, and it uses simulation, machine learning and decentralized optimization and reasoning to help decision making. In the life cycle of MAS development from design to construction to operation, the fidelity of an ICES DT in terms of accurately representing the operations of the eventual "real twin" will increase as more elements of the MAS get installed and provide data streams to the DT. At the very beginning or design stage of a MAS, it is possible to gain high-fidelity results by using data streams from small table-top or small laboratory-scale replicas of the eventual hardware to be installed. Such replicas are often referred to as hardware-in-the-loop (HIL.) If you have seen the movie Apollo 13, you will remember that the astronauts trained on a "digital twin" consisting of a laboratory-scale replica of the command module and simulation software. Besides being a training aide, the Apollo DT uplinked to the mission operations (i.e., the real twin) to compare data between the ideally simulated operations and the actual data streams from the mission. When an explosion destroyed a piece of hardware during the mission, the less-than-ideal operating conditions of the real twin could be replicated on the digital twin to find a solution and save the mission. So, by the operating stage of the MAS life cycle, the DT becomes part of the operating system for the real twin. It is important to note here that SCE began using digital twin technology in 2022 to consider the optimal way to increase customer adoption of DER (Castaneda & Ioan, 2022).

Addressing DER Adoption Issues With Modern ICES

Addressing the first issue of hosting capacity limitations involves energy managers' knowledge of commercially available medium voltage power electronics. These have long been used at customer sites in Europe but have not commonly been used in North America, particularly not with their software capabilities enabled. However, this approach was reviewed and had gained acceptance among SCE engineers in 2012. Addressing the second issue of the difficulty of solving centralized control and optimization algorithms involves energy managers' knowledge and potential use of an open-source decentralized control and optimization approach utilizing a DT platform called Dynamic Monitoring and Decision Support (DyMonDS) invented by MIT Professor Maria Ilic.

The DyMonDS DT platform uses price signals and hierarchical communications to simulate a modern ICES project consisting of customer-owned DER such as on-site power, energy storage and demand-side management organized into the local-area distribution networks (i.e., microgrids) that interconnect with each other, the wide-area utility-owned transmission and distribution network, and certain transportation carriers (i.e., vehicle-to-grid, ship-to-grid and rail-to-grid power).

DyMonDS simplifies the simulation of an ICES global equilibrium for the simultaneous objectives of affordability, reliability, and environmental sustainability by requiring each DER to only communicate their amount and rate of real and reactive power with their most immediate neighbor. Such simulations enable training and cybersecurity exercises during the design phase of the complex system. At this stage, data streams from working DER at the Massachusetts Institute of Technology Lincoln Lab (MIT LL) provide a high degree of fidelity in terms of simulations accurately reflecting actual ICES operations. Thereafter, the DyMonDS DT acts as the operating system for the actual installed ICES (i.e., the real twin) by conducting day ahead simulations, exchanging data with the real twin during real time operations, and making any error corrections between the two.

Addressing the third issue involves energy managers integrating knowledge of power electronics and DT as previously described into a methodology for planning and delivering actual ICES projects. This methodology was inspired by ICES researchers at TU Delft and organized into a four-level engineering and institutional design methodology (Warner, 2023). See Figure 1.

GridLink Engineering and Institutional Design Methodology					
Level		**Engineering Design**		**Institutional Design**	
Access	1A	Engineering Feasibility	Technology Availability & Feasibility	Institutional Feasibility	Customs, Traditions & Norms
			Permit & Regulation Requirements		Legal Feasibility & Government Support
			Utility Interconnection Requirements		Analysis of Utility Competition
			Electrical & Thermal Demand Forecast		Discounted Cash Flow Model
	1B	Systematic Environment	Basic System Architecture	Institutional Environment	Stakeholder Memorandum of Understanding
			Preliminary Design (30%)		Design-Build-Operate Supplier Subcontracts
			Conditional Interconnection Approval		Real Options Model
Responsibility	2	Design Principles	Complete IEEE Standard Design	Governance	Energy Purchase & Use Contracts
			Engineering Procurement & Construction		Construction Loan Agreement
			UL Certification		Dispute Resolution Procedures
			Final Interconnection Approval		Game Theoretic Profit Sharing Model
Control	3	Control Mechanisms	ISO Registration & Bidding Rules	Organization	ISO-ICES Agreements
			O&M Rules & Procedures		Permanent Financing Agreement
			Other Operational Decision Rules		Day-Ahead Optimal Power Flow Model
Operation	4	Operations	Monitoring & Control Systems	Administration	Accounting & Reporting System
			Information & Communications Systems		Membership Meetings & Support
			Asset Management Plan		Error-Corrected Optimal Power Flow Model

Figure 1. GridLink Engineering and Institutional Design Methodology (Pareto Energy, LTD)

This innovative methodology differs from microgrid planning and project development efforts that only address engineering. It includes four stages of simultaneous engineering and institutional design. It has been developed by Pareto Energy as a standard open-source framework of governance contracts, legal enabling, and decision support software whereby an ICES of integrated local-area microgrids, collectively planned and governed by consumers as infrastructural commons, can enjoy fair competition and equitable profit sharing with the wide-area utility-owned grid.

This approach is developed to ensure money is not wasted doing engineering design that has no social or institutional feasibility, nor is money wasted doing institutional feasibility that doesn't have any engineering aspect. The first level deals with access to the microgrid, the utility grid, and to transportation carriers as applicable. The second level is about assigning responsibilities to the agents that manage that, so the cost of benefits is shared in a way that key stakeholders (labor, communities, and users themselves) can agree to and support. The third level is control which deals with setting up the rules before the start of operations to control the power flows (i.e., the optimal power flows within the microgrid). The fourth (and last) is operations. On the left-hand side, the power electronics engineering design is done by IEEE standards with UL certification.

Educating Energy Managers for ICES Project Development

Energy managers must be able to integrate modern ICES knowledge into the overall Shore Energy Program corporate knowledge. The image shown in Figure 2 captures the broad spectrum of challenges energy managers face that are associated with supporting the development and execution of their overall energy program for their respective chains of command. It is a Navy perspective, distilled from the past seven years of energy program management for the Navy’s 70 installations across 10 Navy regional commands.

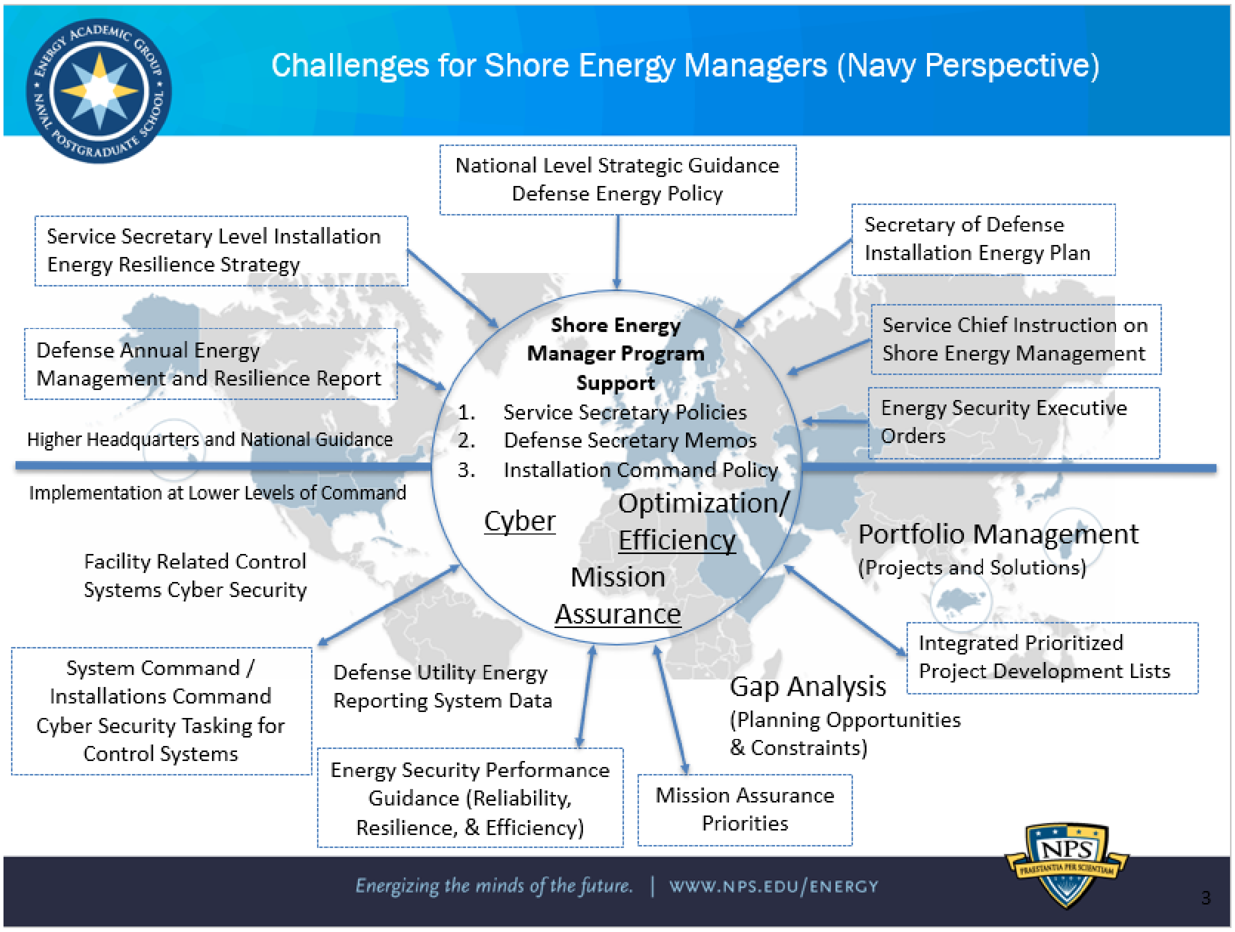


Figure 2. The Challenges Shore Energy Managers Face Managing Their Respective Shore Energy Programs

The challenges shore energy managers face can have both highly technical engineering aspects as well as complex institutional aspects to them. Energy managers therefore need continuous training and education to be able to maintain a common understanding of the current energy program and policies as implemented in their respective services and to maintain knowledge of and proficiency in the use of the latest tools and techniques to aid them in supporting a yearly plan of action and milestones that inform energy project selection. In the Navy, this process consists of four phases aligned with the four quarters of the fiscal year. Phase one is gap analysis across each of the installations that is performed at the installation level with technical assistance (in the case of the Navy) provided by the Navy's facilities engineering Systems-Command (a SYSCOM called NAVFAC).[1] In phase two, alternatives to address the gaps are studied and analyzed. Phase three is where planning and project development occur along with the vetting of different acquisition strategies across the teams, installations, and regions. Other SYSCOMs as well as the Navy regional commands get more involved in this phase as mission assurance is assessed and relative prioritization is worked. The planning and development of viable projects that can really be executed is among the most challenging things an energy manager supports.

Phase four is the coordination of all the energy project requirements developed in phase three, up through the chain of command to Commander, Navy Installations Command (CNIC),

[1] In the Navy, SYSCOMs like NAVFAC are execution agents and *materiel* solution providers for TYCOMs like CNIC that man, train, and equip warfighters.

who is the Chief of Naval Operations (CNO) designated Shore Type-Command (TYCOM) for final review and approval. All energy project submissions from each installation are historically due by September 30 every single year. It's expected that the submissions from each installation are vetted through leadership at the installation level through the installation and regional staff and commands for coordination and prioritization of effort.

Education for shore energy managers must reinforce official Shore TYCOM guidance and process while simultaneously integrating essential knowledge in planning and development of modern ICES projects for defense communities. This can be done through a four-course graduate certificate program for entry to mid-level energy managers. The graduate certificate courses can be organized as follows:

1. Shore TYCOM Energy Requirements Development and Project Feasibility Assessments
2. Analysis, Design Principles, and Governance of Shore TYCOM Energy Projects
3. Planning, Development, Control, and Organizational Considerations of Shore TYCOM Energy Projects
4. Coordination, Execution, Operations, and Administration of Shore TYCOM Energy Projects

A corresponding four-day-long condensed senior managerial level short-course for public works officers, regional energy program managers, and other defense community stakeholders (e.g., union or community leadership, mayoral staff, utility company managers) can also be offered. A Coherent Resilient Tabletop Exercise (CORE TTX) alternative to the short course can also be offered to focus on specific locations, scenarios, and specific energy resilience project planning. The four course certificate, senior manager short course, and CORE TTX can all leverage practical exercises and game-based learning.

The objectives of the Shore TYCOM Energy Certificate Program and Corresponding Senior Manager Short Course are to educate installation and regional energy managers, public works professionals, and other key defense community stakeholders on the key modern energy technologies, systems, and business approaches that are critical to ensuring reliable, affordable, and sustainable power for the shore enterprise, and to certify they have the combination of fundamental engineering and institutional knowledge for effective shore energy planning and management. The coursework and instructional content will be informed by official guidance as promulgated by CNIC, as well as ICES research from the "Applied Microgrid Design and Digital Twins Research and Education for Municipal, Port and Military Energy Planners" effort under the CRADA between NPS and Pareto Energy LTD, led by the Energy Academic Group.

The governance of Shore TYCOM projects in the second course will cover the Installation Energy Program Summary (IEPS); for projects, the Energy Project Selection Process (EPSP); etc. For background, CNIC under the authority as the Shore TYCOM partners with NPS to build out the TYCOM roles and responsibilities, partnerships, and the definition of "manning, training and equipping" energy managers around the globe.

Through the four-course sequence and corresponding short course, the content and instruction are envisioned to foster the knowledge of energy managers during the four phases of installation energy management efforts each year and improve understanding of the simultaneous engineering and institutional aspects of ICES planning and development. An existing case study of the Puerto Rican power grid using DyMonDS, that has shown how Puerto Rico's electricity cost and resilience could be improved through a modern ICES project using key technologies, may be explored in more detail in the four-quarter graduate certificate course. This MIT LL case study is motivated by the recognition that serving highly distributed electric power loads during extreme events requires innovative methods (Ilic et al., 2019). To do this, the type and locations of the most critical equipment, innovative methods, and software for operating the electrical system most effectively must be determined. Existing systems must be both hardened and further enhanced by deploying DER and local reconfigurable microgrids to manage these newly deployed DER. A key learning outcome of the instruction, practical exercise, and game-based learning is achieving a fundamental understanding of how to scope, plan, design, and deliver, operate, and procure power from modern ICES projects in a way that effectively hosts DER for use during both normal operations and grid-failure situations. Traditionally, utility companies rely on excessive amounts of centralized reserve generation to mitigate failures. This increases the cost of normal operations and nullifies the potential of DER to improve reliability by meeting loads during grid failures. Innovative distributed utility business approaches that overcome these limitations are going to be in demand. Innovative public private partnership distributed utilities can use the open source DyMonDS DT platform to do so. Energy managers will need to understand DyMonDS and how to leverage it for training, exercises, and independent government estimation and verification.

Prototype ICES DT Reference Case With SCE

As a first Navy installation DT reference case, Commander Navy Region Southwest (CNRSW) has expressed interest in potential collaboration between NPS and SCE on a prototype DT reference case simulation of a modern ICES to serve NBVC and the Port of Hueneme. This initiative supports the Navy's commitment to developing resilient microgrid projects capable of sustaining operations for at least 21 days during central grid outages, as mandated by the Department of Defense's March 2024 Unified Facilities Criteria for Resilient Installation Microgrid Design. The modern ICES DT will model decentralized control and distributed optimal power flow, enabling microgrids to act as primary power sources for installations and surrounding communities, rather than serving solely as backup systems. By leveraging federal grants, tax credits, and third-party financing, the prospective project would aim to ensure affordability and reliability while fostering energy resilience.

The ICES DT will also serve as a critical tool for workforce development, addressing concerns about skilled labor availability for designing, building, and operating ICES microgrids. It will provide training for electricians, engineers, and optimization specialists, equipping them to meet the needs of critical infrastructure operators. Additionally, the DT will simulate responses to natural disasters and cyber or physical attacks, enhancing preparedness and operational resilience. This effort aligns with SCE's Grid Technology and Innovation (GTI) group's objectives to increase DER adoption and defer transmission and distribution investments.

The project will support the integration of proven technologies, including a power electronic interconnection platform widely used in Europe and the DyMonDS platform, which offers decentralized control and optimization capabilities. Institutional innovations would ensure long-term governance and stakeholder alignment. Union labor pension funds will initially own the ICES and transition ownership to the community through stock purchasing plans. The project would also address regulatory requirements for distributing ICES power to multiple consumers, including obtaining a Certificate of Public Convenience and Necessity (CPCN).

To support education and research, the reference case ICES DT will inform certificate and short course content, fostering knowledge transfer and practical application. Stakeholders will participate in a structured pathway that includes earning a seminar digital badge for foundational training, attending an ICES short course at the CNRSW Broadway Complex for advanced skills development, and/or engaging in a Coherent Resilience Tabletop Exercise (CORE TTX) to apply learned concepts through an ICES planning charrette.

With SCE's support to designate the NBVC DT pilot project as a GTI demonstration initiative, this collaboration will address barriers to DER adoption, secure funding opportunities, and establish a pathway for ICES implementation. By leveraging proven technologies, institutional models, and educational frameworks, the reference case DT pilot project aims to enhance energy resilience, foster innovation, and support the Navy's energy resilience goals.

Prototype ICES Project Delivery

Following a prototype ICES DT reference case, and assuming favorable outcomes, a modern ICES prototype project will aim to enhance energy resilience for both the Port of Hueneme and NBVC. Inspired by Kirk Phillips' (2023) innovative geothermal prototyping approach, this prototype project will leverage Other Transaction Authority (OTA) under 10 U.S.C. § 4022 to bypass lengthy traditional DoD acquisition processes, perhaps enabling other rapid 2-year pilot ICES projects to further de-risk modern ICES implementation. This flexible approach aligns with the DoD's 99.9% resilience mandate and would establish a baseline for other Naval ICES initiatives.

Because the utility macro-grid, responsible for wide-area power transmission and distribution, struggles to meet growing demands for energy resilience and reliability, this leads to vulnerabilities in communities supporting critical infrastructure. In response, state public utility commissions are adopting distributed utility models that can deploy microgrids within 18 months. Federal tax credits and grants covering up to 80% of capital costs make microgrids more affordable and reliable than the macro-grid while offering enhanced resilience against disasters and disruptions.

Through the CRADA led by the EAG at NPS, the prototype project will refine best practices for third-party financing of modern ICES projects to serve defense installations and their surrounding communities. Using OTA, the project will prototype a distributed utility company and a non-Federal Acquisition Regulation contracting vehicle to standardize third-party financing for installation energy infrastructure. The OTA structure will resemble a Utility Energy Service Contract (UESC) and be executed by CRADA partners or incumbent utility companies.

The ICES prototype project will integrate key technologies and innovations, including:

- **Power Electronics Interconnection and Control**: Off-the-shelf power electronics and optimal power flow software to enable reliable transactions between microgrids, the utility macro-grid, and transportation carriers as applicable.
- **Digital Twin Technology**: Model-based design using NPS's test bed and MIT's open-source software for optimization, training, and cybersecurity testing, aligning with DoD energy goals.
- **Institutional Innovations**: A municipal corporation framework for consumer governance of microgrids, supported by optimization software and regulations to ensure equitable benefits and fair competition.
- **Workforce Development**: Labor pension funds will finance microgrids, enabling unions to establish training programs for skilled labor. NPS will collaborate with university think tanks to train engineers and electricians while developing curricula for Naval and other defense personnel.

This modern ICES prototype project will represent a transformative approach to energy resilience, fostering innovation, affordability, and reliability for defense installations and their supporting communities.

Conclusion

Enhancing energy resilience in Navy Region Southwest requires a paradigm shift in how installations and their surrounding defense communities approach energy planning, acquisition, and infrastructure development. The vulnerabilities of aging macro-grids, escalating energy demands, and persistent challenges in DER adoption underscore the urgency for innovative solutions. This paper has demonstrated that modern ICES, supported by technologies such as DT and decentralized optimization platforms, offer a viable pathway to bridge capability and capacity gaps while ensuring operational resilience.

By integrating modern ICES planning and development into the Navy's Shore Energy Program, energy managers can address hosting capacity limitations, overcome centralized control challenges, and foster DER adoption through institutional and technological innovations. The proposed four-level engineering and institutional design methodology provides a structured framework for planning and delivering ICES projects, ensuring alignment with mission-critical requirements and stakeholder priorities. Furthermore, the educational initiatives outlined—ranging from graduate certificate programs to tabletop exercises—will equip energy managers and public works professionals with the knowledge and skills needed to navigate the complexities of modern energy systems.

The prototype ICES DT reference case for NBVC and the Port of Hueneme exemplifies the transformative potential of these systems. By leveraging advanced technologies, institutional models, and innovative financing mechanisms, this initiative aims to enhance energy resilience, reduce reliance on vulnerable macro-grids, and support the Navy's energy security goals. The use of OTA for rapid prototype project delivery further highlights the adaptability and scalability of ICES as a solution for defense energy challenges.

In conclusion, modern ICES represent a critical evolution in energy resilience planning for Naval and other defense installations and their surrounding communities. By fostering collaboration between stakeholders, integrating cutting-edge technologies, and prioritizing education and workforce development, the Navy can achieve secure, reliable, and sustainable energy solutions. This approach not only addresses immediate capability and capacity gaps but also establishes a foundation for long-term resilience and innovation across the shore enterprise.

References

American Society of Civil Engineers. (2025). *2025 infrastructure report card: Energy infrastructure*. https://infrastructurereportcard.org/cat-item/energy-infrastructure/

Carlson, C. (2024, Jan 12). Storm causes millions in damage at the Port of Hueneme; power system for ships shut down. *Ventura County Star.* https://www.proquest.com/newspapers/storm-causes-millions-damage-at-port-hueneme/docview/2913482113/se-2

Castaneda, J., & Ioan, A. (2024). *Electric access system enhancement: Assessment of a distributed energy resource management system for enabling dynamic hosting capacity* (Publication No. CEC-500-2024-064). California Energy Commission.

Cheng, Z., et al. (2018, March 21). To centralize or distribute: A comparison of advanced microgrid management systems. *IEEE Industrial Electronics Magazine*, *12*(1).

Clausewitz, C. V. (2009). *On war: The complete edition*. Wildside Press.

Donnelly, M. K., et al. (1996, May 1). Impacts of the distributed utility on transmission system stability. *IEEE Transactions on Power Systems*, *11*(2).

Environmental Security Technology Certification Program. (2017). *Cost and performance report for Portsmouth Naval Shipyard microgrid and ancillary services – Kittery, ME*.

Ilic, M., Salcedo Ulerio, R. O., Corbett, E. G., Austin, E., Shatz, M., & Limpaecher, E. R. (2019). *A framework for evaluating electric power grid improvements in Puerto Rico*. Massachusetts Institute of Technology. https://www.ll.mit.edu/r-d/publications/framework-evaluating-electric-power-grid-improvements-puerto-rico

Joint Chiefs of Staff. (2020). *Joint planning* (Joint publication 5-0).

Kubli, M. (2023). A typology of business models for energy communities: Current and emerging design options. *Renewable and Sustainable Energy Reviews*, 176.

Lawrence Berkeley National Laboratory. (2012). *An examination of temporal trends in electricity reliability based on reports from U.S. electric utilities*. OSTI. https://www.osti.gov/biblio/1055706

Lovins, A. (2002). *Small is profitable.* Rocky Mountain Institute.

MilitaryBases.com. (n.d.). *California military bases*. Retrieved from https://militarybases.com/california/

National Academies Press. (2012). *Terrorism and the electric power delivery system*. https://nap.nationalacademies.org/read/12050/chapter/5

Ostrom, E. (2005). *Understanding institutional diversity*. Princeton University Press.

Ostrom, E. (2010, October 1). Polycentric systems for coping with collective action and global environmental change. *Global Environmental Change-Human and Policy Dimensions*, *20*(4).

Pareto Energy. (2023). *Proposal for a St Elizabeths microgrid, submitted to the District of Columbia Office of General Services*.

Phillips, K. (2023). Boosting installation energy resilience through geothermal power. In *The Military Engineer, 115*(746), 48–51.

Power Grid International. (2016, August 22). *Protecting aging substations from cyberattacks*. https://www.power-grid.com/executive-insight/protecting-aging-substations-from-cyberattacks/

Pupp, R. (1993). *Distributed utility valuation project monograph*. Electric Power Research Institute.

Reis, I. F. G., et al. (2021). Business models for energy communities: A review of key issues and trends. *Renewable and Sustainable Energy Reviews*, *144*.

Smart Electric Power Alliance. (2024, September 19). *The aging US power grid: Navigating toward modernization*. https://sepapower.org/knowledge/the-aging-us-power-grid-navigating-toward-modernization/

Smeers, Y., et al. (Eds.). (1997). Distributed resources: Toward a new paradigm of the electricity business. *The Energy Journal, Special Issue*.

Southern California Edison. (2023). *Distributed energy resources dynamics integration demonstration* (Project No. SCE GT-18-0019). California Public Utilities Commission

Electric Program Investment Charge Research, Development and Deployment Program. https://database.epicpartnership.org/project/32494

Southern California Edison. (2025, February 12). *Los Angeles Times op-ed*. https://newsroom.edison.com/los-angeles-times-op-ed

Strange, J. L. (1996). *Centers of gravity & critical vulnerabilities: Building on the Clausewitzian foundation so that we can all speak the same language* (2nd ed.). Marine Corps University.

U.S. Department of Energy, Office of Electricity. (2021, October 28). *Addressing security and reliability concerns of large power transformers*. https://www.energy.gov/oe/addressing-security-and-reliability-concerns-large-power-transformers

U.S. Energy Information Administration. (2018, July 20). *Major utilities continue to increase spending on U.S. electric distribution systems*. https://www.eia.gov/todayinenergy/detail.php?id=36675

Wan,Y., et al. (1993). *Factors relevant to utility integration of intermittent renewable energy technologies*. National Renewable Energy Laboratory.

Warner, G. (2023). *Optimized planning and operations of integrated community energy systems: Introducing the dynamic monitoring and decision systems framework (dymonds)* [Lecture]. Naval Postgraduate School, Monterey, CA, United States. https://nps.edu/web/nps-video-portal/-/optimized-planning-and-operations-of-integrated-community-energy-systems-introducing-the-dynamic-monitoring-and-decision-systems-framework-dymonds-

THIS PAGE INTENTIONALLY LEFT BLANK

Productionizing Data Science at Sea Capabilities

John McDonnell—is the PMW150 Technical Director (acting) and supports multiple Future C2 initiatives. While at NIWC PAC he developed Robotic, AI, Virtual Reality, and Advanced C2 capabilities. He currently leads the development of various Battle Management Aids including Weapon-Target Pairing tools, Warfighting Capability Assessments, and Data Centric Security efforts aligned with interoperable Future C2 Coalition objectives. [john.r.mcdonnell12.ctr@us.navy.mil]

Chuck Datte—is the Lead Software Engineer for the Data Science at Sea Software Factory effort. He established the NIWC PAC Rapid Innovation and Test Environment (RITE) and has applied the underlying tenets to multiple Joint and Naval programs including the Air Force JSPOC mission system, MTC2, and Project Overmatch. [charles.datte@sage.northcom.mil**]**

Nick Gizzi—is the PEO C4I/PMW150 APM for S&T. He leads Future C2 initiatives including the development of the Force Composability model operating Left of the Kill Chain and has established a service-mesh architecture that hosts Battle Management Aids as micro-services providing DoN, Joint, and Coalition capabilities. [nicholas.j.gizzi.civ@us.navy.mil]

Abstract

Delivering software capabilities to the operators has never been more challenging, nor has there ever been as much infrastructure and tooling to achieve this objective. This paper outlines an effort to take Fleet-developed operational prototypes and rapidly productionize the capabilities to scale out to afloat and ashore units. The Data Science at Sea initiative is a Fleet effort whereby operators and analysts develop capabilities that are required for recurring operations. The Data Science at Sea Software Factory seeks to establish developer guidelines and a software engineering environment that enable uniformed developers to develop more mature prototypes that can be rapidly containerized, ensure cyber compliance, and provide availability for deployment to any Naval unit that desires the Fleet-developed capability. The Data Science at Sea initiative requires sustainment to provide continued prototyping and production support to proliferate and scale applications to the broader Naval warfighting community.

Introduction

The Data Science at Sea (DS@S) initiative is a Fleet effort whereby operators and analysts develop capabilities that are required for operations and have not been provided in a sustainable manner from the acquisition community. The DS@S initiative kicked off in 2021 on the Carl Vinson (CVN-70) in support of CSG-1 operations. A mixed uniform and civilian data science team developed operational capabilities that have yielded operator time savings and provided new ways to analyze and display data. These operator-developed and informed tools have not had to go through a protracted requirements and POM process to field and can proliferate organically to establish wide operator adoption. The DS@S is warfighting capability innovation at the edge. The DS@S Software Factory effort is focused on supporting the proliferation of these capabilities to other CSGs and warfighting units where the manpower may not be as trained or equipped to take advantage of DS@S innovation.

Delivering software to the operators has never been more challenging nor has there ever been as much infrastructure and tooling to achieve this objective. This paper outlines an effort to take Fleet-developed operational prototypes and rapidly productionize the capabilities to scale out to afloat and ashore units. The DS@S initiative is a Fleet effort whereby operators and analysts develop capabilities that are required for recurring operations.

The overarching objectives of the DS@S initiative include man, train, and equip DS@S teams based on Naval Information Forces (NAVIFOR) Requirements and resource the “Uniform Digital Talent” to ensure that there is a pipeline of personnel. Secondary to the Fleet-driven need to develop required capabilities in-situ is converting the DS@S operation prototypes to containerized applications that can be deployed across the Fleet from the Project Overmatch Application Arsenal.

There are currently four organizations actively supporting the Fleet's DS@S initiative:

- Center for Naval Analysis (CNA): Provides personnel to support DS@S tools, and supports prototype development and operation.
- Office of Naval Research (ONR) Tech Solutions/PMW150: Supports the on boarding of DS@S capabilities to the DS@S Software Factory and supports the software engineering from prototypes to scale capability releases to Naval units.
- Project Overmatch/Warfighting Data Services (WDS): Provides the DS@S Toolkit and personnel to support the development and implementation of operational prototypes.
- STRATCOM Global Data Initiative (GDI)/ACE-M: Provides GUNNS/ACE-M platform for running DS@S tools and applications and curates data that is disseminated via broadcast.

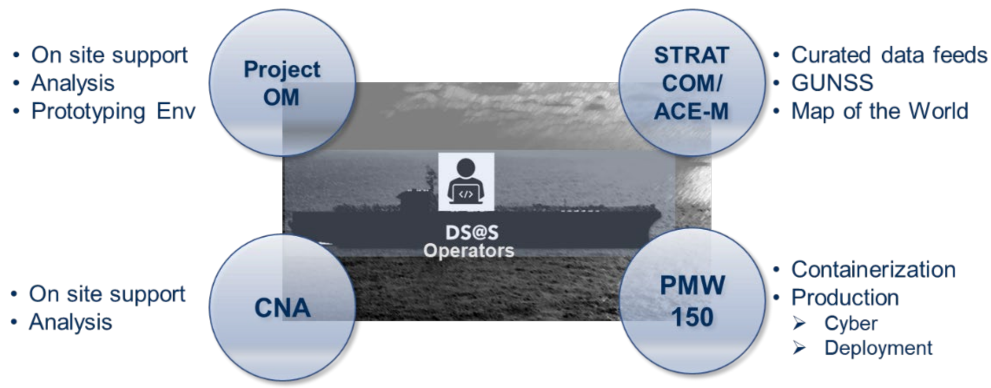


Figure 1. Organizations Supporting the DS@S Initiative

The DS@S-support teams are depicted in Figure 1 in conjunction with the functions that each provides. A biweekly stakeholder update has been established to articulate the status of the ONR Tech Solutions Topic 1000, Productionizing DS@S Capabilities. In addition to conveying status, the biweekly session facilities cross-organization collaboration to minimize redundancies across the efforts as well as establish synergies across the teams moving forward.

DS@S Software Factory Objectives

The purpose of the DS@S Software Factory is to facilitate converting Fleet-developed prototypes to cyber-secure, containerized applications that can be put through a production software engineering environment such as the Overmatch Software Armory (OSA) and deployed via the Overmatch Application Arsenal.

A software factory is a structured environment that uses standardized tools, processes, and reusable components to accelerate and improve software development, aiming for efficiency and quality through automation and assembly-line techniques. A software factory should reflect the desired process, incorporate recommended tools to implement the process, and host those tools in an environment that is both cost-effective and accessible to software developers, testers, and integrators.

A software factory supports the following concepts:

- Structured Approach: Software factories adopt a systematic approach to software development, mirroring manufacturing processes with defined workflows and standardized practices.
- Standardized Tools and Processes: Leverage a collection of tools, templates, and methodologies to streamline the development life cycle, from requirements gathering to deployment.
- Reusable Components: Software factories emphasize the reuse of code, modules, and other assets to reduce development time and effort.
- Automation: Automation is a key aspect of software factories, automating tasks like testing, deployment, and code generation to increase speed and reduce errors.
- Continuous Delivery: Software factories are often designed for continuous integration and delivery (CI/CD), allowing for frequent releases and updates.
- DevSecOps and Agile Principles: Software factories are often rooted in DevSecOps and agile software development principles, promoting collaboration, feedback, and iterative development.

The benefits of software factories can lead to increased efficiency, higher quality software, faster delivery times, and reduced costs.

Productionizing DS@S Pilot

The ONR Tech Solutions Office has initiated a Productionizing DS@S Capabilities Topic at Fleet request. This proof-of-concept pilot works with DS@S Officers-in-Charge (OICs) to prioritize the most mature and impactful DS@S operational prototypes. The DS@S OIC prioritized capabilities included Pelican which supports geospatial correlation for weapon-target pairing, and HORUS which supports the visualization of find, fix, track, target, engage, and assess (F2T2EA) kill webs.

The ONR Tech Solutions effort was broken down into three distinct phases to establish a pilot DS@S Software Factory for productizing DS@S capabilities. These phases are shown in Figure 2 and described by:

- Phase I:
  - Develop SOPs for containerizing and ensuring cyber compliance for DS@S capabilities
  - Demonstrate the SOP for select DS@S applications
  - Improve workflow automation and data ingest
- Phase II:
  - Demonstrate interoperability with the MTC2 Tactical Planning Tool and the CANES/ACS CJMTK Common Mapping Service
- Phase III:
  - Productionize capabilities through the OSA
  - Release capabilities to the Fleet via OSA's App Arsenal

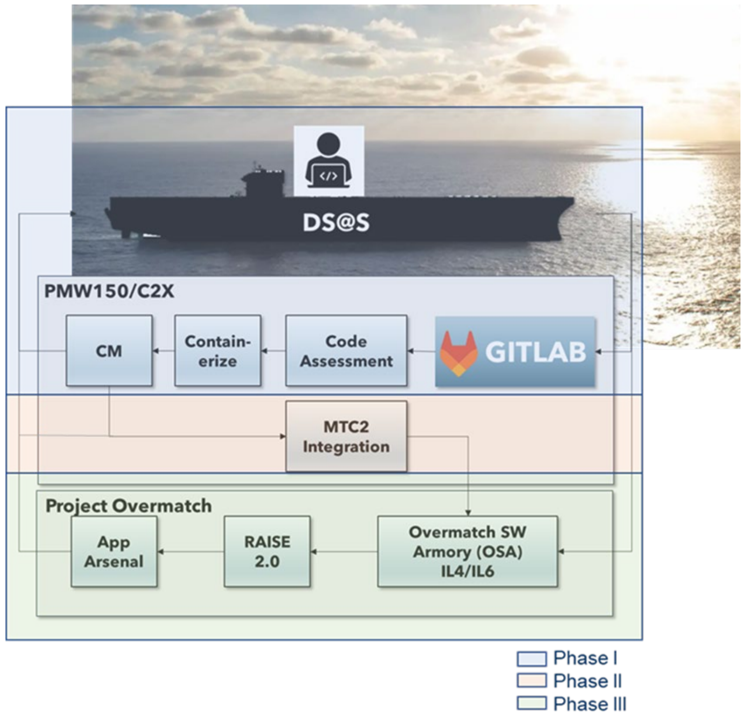


Figure 2. Phases for ProductionizingDS@S Capabilities

During Phase I of this effort a digital twin of the operational environment has been established in the PMW150 Command and Control Experimentation (C2X) facility. The digital twin reflects the capabilities that are included in the DS@S Toolkit and includes the ERSI ArcPro environment for running DS@S geospatial tools and JUYPTER for running the python implemented prototypes that have been developed. The digital twin also supports the STRATCOM ACE-M baseline including the NGA Map of the World that is used in the deployed environment. As part of the Phase I objectives, DS@S capabilities are assessed to decouple data from the code base and determine where automation can be incorporated into the prototype to improve the workflow. By decoupling sensitive data from the code base the code can be transferred via a Data Transfer Request (DTR) to the IL4 development environment so that a broader set of developers can have access to mature the capability.

Table 1 lists the components of the digital twin along with other software that reflects the run-time time environment for geospatial analytics.

Table 1. The Digital Twin Along with Supporting Run-Time Tools

VM	OS	Software	Version
C2XACAS	RHEL7	Nessus	8.13.1
c2xansible8yum	RHEL8	nginx	1.14.1
c2xarcgis	RHEL8	ArcGIS Enterprise	10.9.1
c2xacem	CentOS 8	ACE-M	2.X
		Map of the World	3.68.1
		Khonsu (Python)	1.0.7
dsas-jupyternb	W10	Python	3.12.4
		JupyterLab	4.2.4
		ArcGIS Pro	3.2.0
		Khonsu (ArcGIS Pro)	2/18/2025
		Horus	1/7/2025
		Pelican	1/7/2025
		Sahara	1/7/2025

The objective of Phase II is to demonstrate interoperability between select DS@S-developed capabilities and PMW150 PoRs and Project efforts. For example, synergies can be realized between the PMW150 Maritime Tactical C2 (MTC2) Tactical Planning Tool (TPT) with DS@S geospatial analytics to convey a spatial-temporal planning model on single pane of glass. These capabilities have been submitted as part of the FY27 Program Objective Memorandum (POM27).

The focus of Phase III efforts is to put select DS@S capabilities through the OSA software engineering production environment. The MTC2 PoR Leads are facilitating putting select DS@S capabilities through OSA as a MTC2 sub project. This includes meeting the MTC2 PoR on-ramp requirements.

DS@S Software Factory Process

The DS@S Software Factory process was developed to take sailor and analyst developed operational prototypes and provide software engineering to mature them as candidate capabilities to take into a production environment. The discrete steps are readily incorporated into a dashboard reflecting the status for each DS@S-developed capability. The current focus is to provide capabilities that are fielded in an afloat environment and hosted by CANES/Agile Core Services (ACS). The process can be easily tailored for other environments including shore-hosting and other Programs of Record (PoRs) production environments.

The current DS@S Software Factory process is delineated by:

1. Sailors and analysts develop capabilities afloat
2. Capability is posted to Collaboration at Sea as an agreed-upon communication channel for supporting Fleet submissions

3. Program or Project Transition team downloads from Collaboration at Sea to the digital twin intermediary staging environment
4. DS@S software is installed on the digital twin, evaluated against a set of submission checklists and evaluation reports. A submission report is generated, and the code is checked into CM
5. Refactor and rearchitect code as needed. Run validation testing and update CM
6. Post artifacts including Code, Documentation, Test Reports, and Briefs to the Navy Lift DS@S Collaboration site
7. Fleet validation—receive operator feedback via an on-prem instance
8. Incorporate Fleet updates
9. Containerize, scan, and test to align with target PoR. Update CM
10. Conduct OSA Onboarding TEM
11. Instantiate OSA project with the containerized capability
12. Develop CI/CD scripts and customize tooling
13. Execute the RAISE 2.0 process
14. Implementation on OpenShift, service mesh and ACS integration
15. Expose DS@S containerized capability in Application Arsenal (AA)
16. Operators pull DS@S capability from AA to install

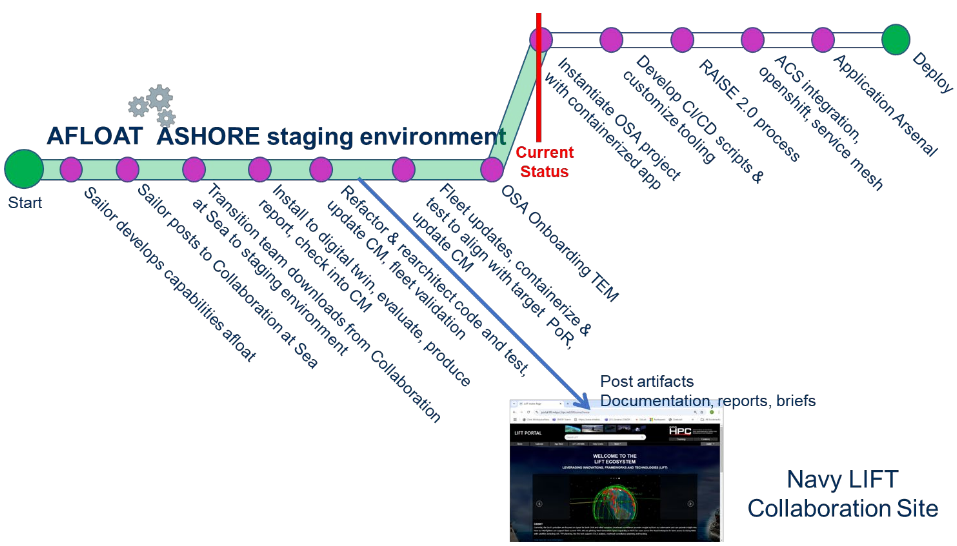


Figure 3. The DS@S Software Factory Process

DS@S Software Factory Implementation

The DS@S Software Factory has been implemented as a combination of the DS@S Toolkit, the GUNSS/ACE-M platform, a Test Manager tool, a Test Deployment environment, and leverages the Navy Lift collaboration software engineering environment. These components are shown DS@S Software Factory box in Figure 4.

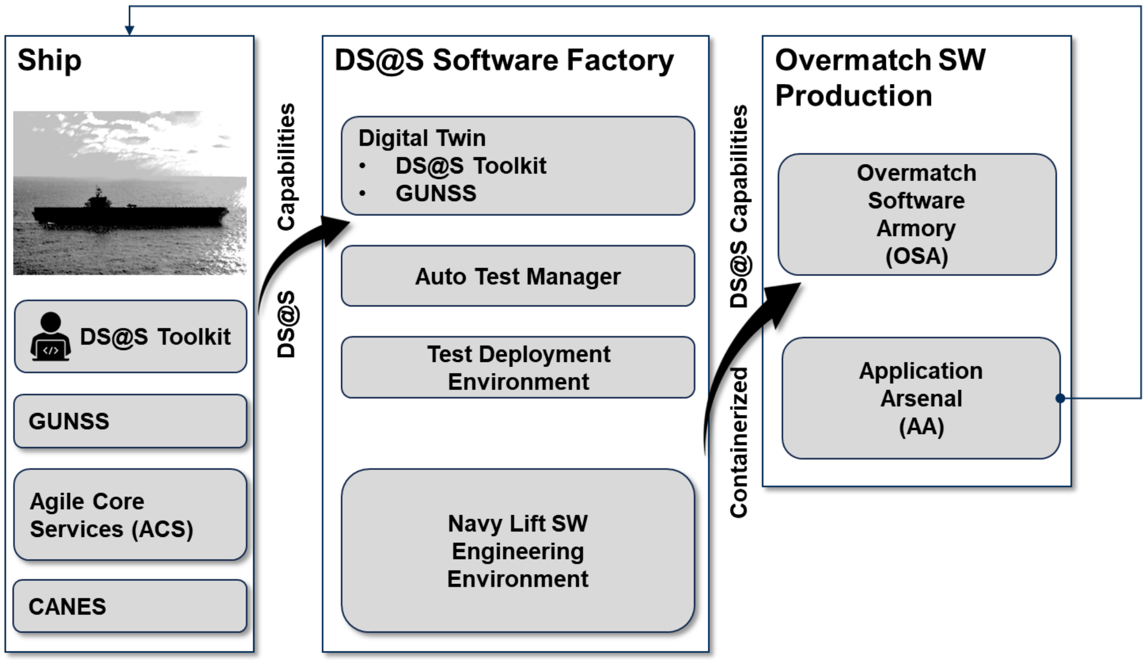


Figure4. The DS@S Software Factory Components and Workflow

It is noted that that this pilot project sought out a software engineering environment that is cost-effective, has a low barrier to entry for participating developers, and provides collaboration tools. For the current effort, Navy Lift meets most of the desired criteria by supporting common toolsets used in other DevSecOps environments.

Navy Lift provides a set of software engineering tools including Bitbucket, Artifactory, and Jenkins to support continuous integration and continuous deployment (CI/CD). It also provides SonarQube for static code analysis. Other tools for conducting container and Web-application vulnerability scans are being explored as potential add-ins for Lift or implementation in the DS@S Software Factory lab environment as an interim solution. These products are implemented in a manner to align to the Project OSA software engineering pipeline.

Once the DS@S Software Factory process is complete, DS@S-developed capabilities are delivered to a production environment. For deployed applications that run on CANES/ACS, OSA provides a Rapid Assess and Incorporate Software Engineering (RAISE) Platform of Choice (RPOC) to ensure cyber compliance by inheriting the security controls of the platform that has an Authorization to Operate (ATO).

An automated Test Manager (TM) is incorporated into the DS@S Software Factory to support the conversion of Fleet-developed prototypes into cyber-secure, containerized applications. Its automated test execution capabilities ensure that these prototypes are thoroughly tested for security and functionality before being integrated into production environments like the OSA. The TM leverages a comprehensive suite of technologies and tools, including GUI capture and playback, message generation and reception, scenario control, reporting, and traceability. These standardized tools streamline the testing life cycle, from requirements gathering to deployment. The TM emphasizes the reuse of test scripts, modules, and other assets, reducing development time and effort. This aligns with the software factory's goal of efficiency through reusable components. Automation of tasks like test execution, reporting, and traceability increases speed, reduces errors, and supports continuous testing. The automated TM provides increased efficiency via repeatable processes resulting in higher quality software and reduced delivery times. The TM is provided to the DS@S Software Factory under a cost-free license model and is implemented in the on-prem environment.

Conclusion and Next Steps

DS@S operational prototyping is the epitome of speed to capability and provides immediate utility to Fleet operational needs. Productionizing capabilities ensures a cyber compliant application posture while scaling to Naval platforms. The approach and methods outlined in Productizing DS@S capabilities is aligned with the USN Information Superiority Vision to innovate and scale as well as follow modern development practices to deliver operational capabilities at speed and scale via DevSecOps practices with integrated security measures at the early phases of the software development life cycle.

There is a need to grow the cadre of Uniform and in-situ developers. The DS@S Production environment has been selected by the Naval Postgraduate School (NPS) as an OPNAV-submitted research topic for the 2025/26 academic year. Other avenues being investigated include collaboration with emerging Naval Software Factories. These collaborations could converge software engineering pipelines and practices in conjunction with other Sailor (and Marine) developers. Finally, the DS@S tenets could be followed to support establishing a Data Science at the MOC (DS@M) initiative.

Both Uniform-driven development and the resulting Fleet-developed warfighting products require sustainment to provide continued prototyping and production support to proliferate and scale applications to the broader Naval warfighting community. While sustainment of select capabilities has been submitted for POM consideration, the sustainment of the DS@S baseline operational prototyping capability and production pipeline requires dedicated initiative.

Acknowledgements

This work was sponsored by ONR Tech Solutions under Topic 1000. The DS@S Production Pilot development and engineering team included Nick Abuzalaf, Jesse Chen, Doug Miller, Zach Vaughter, Nick Keele and Roger Kelly who contributed to this effort.

References

Wien & Nelson. (2023, February). Sailors must do data science at sea. *U.S. Naval Institute Proceedings 149/2/1,440*.

Halton & Reiffen. (2024, February). Lessons learned from year one of data science at sea. *U.S. Naval Institute Proceedings 150/5/1,455*.

Sardiello et al. (2024, June). The case for data science at sea. *U.S. Naval Institute Proceedings 150/6/1,456.*

Department of Navy. (2024, August). *Information superiority vision 2.0*.

Raise Working Group. (2022, October). *Rapid assess and incorporate software engineering (RAISE) 2.0 implementation guide, version 1.0.*

Exploring Factors Impacting Obsolescence Risk in Aerospace and Defense

Lauren Griffiths—is an accomplished procurement and supply chain leader with almost 20 years in the aerospace and defense industry specializing in global procurement, subcontracts, and supply chain. Griffiths's range of experience includes obsolescence management, international supply chain, risk management, and supplier recovery. Griffiths excels in leading, fostering, and empowering team members to develop innovative and creative solutions. She has a strong track record of driving strategic initiatives, improving operational efficiency, and ensuring successful outcomes across global supply chains. Griffiths holds an MBA, PMP, CPCM, and will graduate from Marymount University with a Doctorate in Business Administration in May 2025. [l0g77691@marymount.edu]

Abstract

This paper explores the impact of obsolescence in the aerospace and defense (A&D) industry, focusing on factors driving the escalation of costs associated with obsolescence mitigation and management. Through a focus group study with A&D professionals, themes emerged, including the costs of redesigns, last-time buys, and securing supply. Participants emphasized the need for proactive planning, risk mitigation, and cross-functional collaboration to manage obsolescence. These findings highlight the complexity of obsolescence and its financial impact on an organization. The results offer valuable insights into improving obsolescence strategies and set the stage for further research on the topic.

Introduction

Obsolescence is the problem of not being able to procure raw material, a component, or a major assembly (product) because the original equipment manufacturer (OEM) and distributors no longer manufacture and distribute the product, or it has been replaced in the market by a newer product (Rojo, 2010). The result of obsolescence can be a shortage in availability to meet current and future demand of the market, which can result in the stopping of a production line and contract fulfillment (Sandborn, 2013).

Key products of the aerospace and defense (A&D) industry typically have long life cycle phases from development through sustainment, and higher risk for obsolescence issues (Del Campo et al., 2022). These products encompass large military systems, aircraft, ships, ground-based vehicles, industrial and medical equipment, and electronics. For this industry, which includes the Department of Defense (DoD), foreign military customers, and A&D companies, unplanned obsolescence issues can become expensive and difficult to identify, avoid, mitigate, and resolve.

The problem of obsolescence can become complicated as multiple products and program areas share materials and parts. In avionics, where components are often shared across multiple aircraft types, obsolescence and parts unavailability can significantly drive up costs (Sandborn et al., 2008). This necessitates a thorough analysis of the root causes of high ownership costs for each aircraft type to mitigate the broader financial impact.

In addition to the high costs related to resolving issues of obsolescence, the consequences of not resolving problems of obsolescence for the A&D industry stakeholders means that materials, components, and major assemblies are at a substantial risk of being unavailable to support requirements (DoD, 2024). This means a soldier or sailor is left vulnerable in the field, inflated costs are required to expedite the resolution of the part, and loss of revenue to the business with higher cost to the taxpayer are a result (DoD, 2024).

Problem Statement

Unplanned parts obsolescence in the A&D industry require significant cost and time to resolve. This causes the unavailability of materials, components, and major assemblies to support contractual requirements and customer needs. The impact of obsolescence can be felt from the individual to entire nations. For defense products, a soldier or sailor is left without a major product or weapons system in the field, an expensive solution is required solve the shortage, the product cost significantly increases, and added costs are passed on to the taxpayer (DoD, 2024). Similarly, for commercial aerospace, revenues to businesses decreases, and costs are passed on to the consumer through airfare increases or other methods (Del Campo et al., 2022).

Statement of Purpose

The purpose of this research is to broadly explore the experiences of A&D industry practitioners managing issues of obsolescence. This research will provide valuable insight into the practitioners' perspectives of the issue of obsolescence, to understand the factors that contribute to increased costs associated with obsolescence mitigation, management, and risk realization.

Significance

The significance of this paper underscores the critical need to address obsolescence challenges in the A&D industry. The insights from this paper are valuable for researchers and practitioners in procurement and supply chain within the A&D industry. Furthermore, these findings can influence A&D procurement and supply chain leaders in making strategic decisions to mitigate and manage obsolescence challenges.

Review of the Literature

A literature review of A&D industry obsolescence was conducted to explore the challenges of the issue, strategies to mitigate and manage the risk, and gaps in obsolescence frameworks to identify the factors that result in key cost drivers. Recurring themes of this literature includes proactive versus reactive response, A&D cross industry participation, misconceptions of the problem of obsolescence, significant cost and time to resolve obsolescence, technology and product life cycle, and product and contractual requirements.

Obsolescence

Obsolescence in A&D refers to the critical components, materials, or technologies becoming unavailable, unsupported, or non-compliant due to various factors (Rojo, 2010). Obsolescence can occur due to a number of reasons including technological, industry driven, and regulatory (Bartels et al., 2012). Technological obsolescence occurs when newer advancements result in older technologies becoming outdated (Bartels et al., 2012). Industry or market-driven obsolescence occurs when a manufacturer discontinues a product line due to shifts in demand from their customers or reduced profitability (Bartels et al., 2012). Regulatory obsolescence occurs when there are changes in operational, safety, or environmental requirements that make the incumbent material, component, or technology non-compliant to meet requirements (Bartels et al., 2012). These classifications highlight the complex challenges in sustaining long-life defense systems while keeping up with rapid technological evolution in A&D.

Causes and Challenges in A&D Obsolescence

The issue of obsolescence in A&D occurs due to the rapid advances in technology of the industry's products, a limited and shrinking supplier base, and the extended life cycles of A&D products (Sandborn, 2013).

Electronic components which are heavily used across the A&D industry are affected by the concept known as Moore's Law. Moore's Law states that the number of transistors on a microchip will double approximately every two years, significantly impacting A&D and the acceleration of obsolescence in electronic components (Schaller, 1997). This short rapid technological progression significantly impacts A&D by accelerating the obsolescence of electronic components, as commercial manufacturers phase out older technologies in favor of newer, more efficient ones (Sandborn, 2013). Defense systems, which are designed for multiple decades of operation, are difficult to manage and keep pace with these advancements, leading to increased sustainment costs and the need for costly system redesigns (Bartels et al, 2012). For example, according to the Electronic Industries Alliance (EIA) an average cost for redesigning electronic components and parts can fall between $26,000 and $2 million dollars (Ozkan & Bulkan, 2016). In addition, it is estimated that 3% of the world's electronic parts become obsolete every month (Jennings et. al., 2016). Furthermore, reliance on legacy components creates supply chain vulnerabilities, as discontinued parts become scarce and harder to source, resulting in issues of obsolescence (Solomon et al., 2000). This cycle often results in the need for frequent and costly redesigns or last-time buys (LTB) to maintain product and system operation and functionality (Schaller, 1997).

The problem of obsolescence is not only related to electronics components and products though. While electronic components have long been susceptible to rapid obsolescence due to fast changing and advancing technology, nonelectronic parts such as mechanical components, materials, and textiles are also increasingly facing the same types of challenge (Howard, 2022). This issue occurs in systems and products of the A&D industry such as aircraft, ships, and tanks that have been operating for more than 20 years and are expected to remain in service for multiple more decades. In avionics, as components, parts, products, and systems are phased out and unavailable due to obsolescence, the costs for maintenance, concerns for safety parts, and risk for operational inefficiencies increases (FAA, 2015). The extended life cycle and service life of these products and systems results in obsolescence issues and require proactive strategies for minimize risk of shortages, ensure stability of supply, readiness, and supportability (Howard, 2022).

In addition, regulatory changes and geopolitical factors can increase the instance of or worsen obsolescence issues as materials or suppliers become non-compliant for use (Bartels et al., 2012). Regulatory and compliance requirements determined both product and contractual specifications that suppliers must meet, influencing component availability and sustainment ability (Blanchard, 2004). Environmental and safety regulations, such as Restriction of Hazardous Substances (RoHS) and Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH), can lead to obsolescence when older materials or components are banned, requiring expensive redesigns (Teixeira et al., 2017). Additionally, defense procurement policies often have specific requirements for parts sourcing and life cycle management, which can further worsen obsolescence risks by limiting the flexibility in selecting alternative suppliers or parts (DoD, 2021).

Another complication in A&D obsolescence risk is the use of the same parts and components across multiple platforms and systems. Component reuse across multiple products is a widely adopted strategy to leverage economies of scale within a product family. While this may reduce initial costs and streamline procurement efforts, it can also lead to significant obsolescence challenges. When obsolescence strikes, the anticipated savings can be significantly diminished, as supply chain disruptions impact multiple products simultaneously. The competition for limited resources to resolve these issues can further escalate costs and delays. When a component becomes obsolete or discontinued, it affects all systems relying on that part, worsening supply chain disruptions and increasing sustainment costs for multiple platforms (Bartels et al., 2012). Additionally, the widespread use of common components can result in a single point of failure, making it more difficult to find replacement components or parts and maintain system and product integrity over time (Solomon et al., 2000). To mitigate these risks, organizations must maintain visibility across multiple programs and act decisively to address obsolescence challenges before they compound (FAA, 2015).

In the instance of a business or customer having responsibility to build, deliver, or procure multiple products with the same obsolete component issue compounds the problem. When an unplanned obsolescence issue arises, it can result in a negative impact to a program production line as functional stakeholders such as engineering and supply chain personnel work to resolve the issue. In the meantime, a potentially shutdown production line results in unplanned costs incurred for re-planning and resolving the issue, the program failing to meet its contractual obligations, and the customer not receiving a critical delivery on time.

In addition, technological advancements, particularly in electronics, drive rapid innovation cycles that significantly impact component obsolescence. Moore's Law, which predicts the doubling of transistor density approximately every two years, plays a crucial role in determining the life cycle of electronic components (Schaller, 1997). This rapid evolution leads to shorter product life cycles, forcing industries to manage obsolescence proactively through predictive forecasting, last-time buys, and modular design approaches (Solomon et al., 2000). In A&D, where systems must remain operational for decades, the mismatch between commercial electronic advancements and defense acquisition cycles exacerbates obsolescence challenges. Addressing this issue requires an integrated approach that combines life cycle costing with technology refresh strategies to ensure long-term sustainment without excessive redesign costs (Bartels et al., 2012).

Obsolescence Management Standards, Frameworks, and Gaps Identified

Obsolescence management standards and frameworks provide structured approaches to mitigating risks associated with aging components in A&D systems. When parts or materials become obsolete or unavailable, mission-critical systems can suffer. So these standards and frameworks are important to support readiness and reduce risk of schedule delays, and cost increases by promoting best practices to address potential obsolescence issues before they occur. Many of these standards have been established to inform and guide organizations in both domestic and international, as well as government and commercial, applications in A&D. Below are key standards related to obsolescence management with gaps identified.

Though the obsolescence management frameworks do provide structured approaches, there are gaps that remain in procurement flexibility, supply chain risk mitigation, and cost impact assessment. Addressing these issues requires enhanced integration of predictive analytics, agile and flexible contracting, and comprehensive cost models that support long-term sustainment strategies in A&D.

Despite the scope and coverage of these frameworks, there are critical gaps in addressing procurement challenges, supply chain vulnerabilities, and cost impacts in A&D obsolescence (Chellin & Gallegos, 2024). The gaps discussed underscore the need for continued improvement in standards and frameworks to support management of obsolescence in the A&D industry.

Other Approaches to Managing Obsolescence

Proactive versus Reactive Response

One recurring theme found in literature is the proactive instead of reactive approach and methodologies of obsolete component management as identified by Chellin and Miller (2023) and English (2022). Sub themes of this aspect include part prioritization, mitigating risk, planning for costs, complexity of obsolescence, and planning for technology inserts and upgrades. In addition to the DoD standards and guidelines mandate for proactive management, scholars also advise proactive management of obsolescence.

Proactive mitigation strategies for obsolescence fall primarily into three main categories: supply chain, designing for obsolescence, and planning strategies (English, 2022). Supply chain strategies involve actions to mitigate risk in supply and secure inventory for a product or system life cycle which is accomplished within the supply base and through supplier partnerships. Designing for obsolescence puts emphasis on engineering and technical efforts to create adaptable and reproducible designs that prioritize short term needs with allowance for part alternates, changes, and upgrades in the future to mitigate short- and longer-term obsolescence risks (Chellin & Miller, 2023). Planning strategies include proactive methods such as monitoring and forecasting for obsolescence, technology roadmaps, product and system refresh plans, and developing obsolescence management plans (Bartels et al., 2012).

Alternatively, there are implications for a reactive approach to obsolescence. English (2022) concurred that components found to be obsolete without proactive planning can result in additional complexities for resolution. In addition to the increased costs and time constraints from lack of planning, reactive strategies can result in part unavailability from the original manufacturer which can increase the risk of receiving counterfeit parts from brokers (English, 2022).

Cross Industry Participation

Another focus for literature on obsolescence is cross industry participation. The sub theme of this area includes partnership and coordination among DoD divisions, DAU, A&D corporations, original equipment manufacturers (OEMs) and distributors, and other stakeholders across A&D and commercial industries.

Specialized Partnerships Across Industry

Industry-wide participation has been recently implemented in the DoD's collaboration with the commercial industry for microelectronics. The DoD (2022) reports that the Creating Helpful Incentives to Produce Semiconductors (CHIPS) for American Defense Fund was created to fund the work of the Microelectronics Commons. The Commons purpose is to create partnerships in the industry to strengthen the pipeline of workforce talent, expand capabilities, and further technological advancements in the microelectronics industry in the United States (DoD, 2022). Efforts such as this will bridge the gap of obsolescence and availability within the US market for microelectronics.

Collaboration with Sub-tier Suppliers

English (2022) states that while consolidation occurs at the sub-tier supplier level, which reduces parties involved, interaction and proactive collaboration must be in place to ensure that visibility into obsolescence risks are flowed through the supply chain. For example, the U.S. Army Precision Guided Missiles program has seen consolidation of their supply base of rocket motors down to Aerojet Rocketdyne and Northrop Grumman. Without this open dialogue, DoD and A&D industry customers and suppliers are unable to engage in a proactive response that mitigates risk and impact (English, 2022).

Planning for Costs

Chellin and Miller (2023) conducted interviews among government civilian and defense industry partner employees to obtain insight from industry practitioners experience with DMSMS challenges. They found that among the benefits of a proactive approach to DMSMS and obsolescence are long-term affordability, availability of parts, reduction in schedule lead times, and better maintainability in comparison to reactive responses to the issue. The authors also noted that it is critical to fund these activities before there is an issue to ensure teams can work proactively to mitigate and avoid problems (Chellin & Miller, 2023).

Common Misconceptions about Obsolescence

The third theme identified in literature is that there are common misconceptions or myths surrounding how and when obsolescence planning, mitigation, and realization strategies are needed. This includes the belief that obsolescence is a problem, and older products do not require any obsolescence consideration.

Obsolescence is a Problem to Fix if it Occurs

The first subtheme in common myths is that obsolescence is a problem to fix if it occurs. Obsolescence is an inevitability. As technology cycles continue to accelerate, their component cycles have a shorter and shorter life cycle. With the addition of the recent supply chain constraints, a raw material and semiconductor component shortage has resulted for the A&D industry. Along with the high demand for engineering resources programs are experiencing multiple obsolescence issues and cycles at a time (Del Campo et al., 2022). Proactive instead of reactive strategies to obsolescence management must be planned to ensure continuity for a program's ability to maintain delivery obligations to its customers. In addition, systems utilizing electronic parts may be more likely to see DMSMS and obsolescence issues resulting in significant costs and time to resolve.

According to English (2022), obsolescence of electronic components, with microcircuits, is almost always inevitable. In the mid 1990s as the DoD moved to commercial-off-the-shelf (COTS) over military specified components, the shelf life of those components decreased dramatically. This is due to the commercial microcircuit product life cycle rate of 18 to 24 months. With the life cycle of A&D system products having a much longer period, the incidence of an obsolescent component issue is much greater (English, 2022).

Older Products Have No Demand

The second subtheme is that older products have no demand, which is inaccurate. Defense programs, products, subassemblies, parts, components, and material are all used well past what is considered a typical life cycle of a product. In the instance war time, critical missions, or even budget cuts, older equipment is used and must be working properly to support defense and the soldier or sailor. Koczanski states that by not utilizing the DMSMS management and planning practices a program will experience negative impacts to their cost, schedule, and system readiness. Programs must be proactive in their management of DMSMS and obsolescence issues to be prepared when one occurs (Koczanski, 2014). As Porter and Plotkin note, the B-52, Phalanx, and THAAD Missile are examples of defense programs and products that are mission critical for the lives of the servicemember and civilians and are well beyond a standard product life cycle (Porter & Plotkin, 2013).

Summary

To summarize, the prior research has focused across four major research themes in A&D obsolescence including proactive management, financial impact, supply chain, and strategic forecasting. Prior research has also included understanding the factors that lead to obsolescence and mitigation, where the focus of cost drivers has not been prioritized. This research will differ from the previous studies in that the objective will be to bridge the gaps in understanding the specific factors of costs that result in high dollar obsolescence issues.

Research Question

After identification of a gap in the literature on obsolescence in the A&D industry, the following research question is posed for this study: What factors drive an increase in the cost of A&D obsolescence mitigation, management, and risk realization?

Research Method: Qualitative Study Focus Group

The literature review provided an overview of the current knowledge and prior research on the topic of obsolescence. However, there were still gaps and ambiguities on the issues of obsolescence from the perspective of subject matter experts in this field. For example, timely data and scholarly research on obsolescence costs across all product areas of the aerospace and industry was not covered. The researcher determined that a guided question and discussion about experiences with obsolescence costs with a group of experienced practitioners could reveal nuances, ambiguities, and enhance the validity and relevance of this research.

Sample

The researcher recruited eight individuals through purposeful selection from the researcher's own network. An additional two individuals were identified through other participants and invited to participate in the focus group study. Individuals identified were noted to have varying levels of direct experience in obsolescence management in the A&D industry.

Data Collection and Analysis

The focus group was conducted with seven participants of the original ten invited. Each participant had varied experiences in managing obsolescence issues. Participants had experience in three different companies in the A&D industry. Participants' experience included one with less than 10 years of experience, three with more than 10 years of experience, two with more than 20 years of experience, and one participant with more than 40 years of experience. Appendix A provides the details of the focus group participants roles, years of experience and type of experience in obsolescence.

Participants were asked six questions which are noted in Appendix B to support the research question, "What factors drive an increase in the cost of A&D obsolescence mitigation, management, and risk realization?" The focus group lasted approximately one hour and generated a video recording, audio recording, and transcript of the session.

Through the small focus group with predetermined questions, participants had the opportunity to share their experiences, perspectives, and valuable insights into their roles in risk mitigation, management, and resolution of obsolescence material issues. Creswell and Poth note that focus groups can be beneficial when the interviewees have similar backgrounds, experiences, and can interact with each other (Creswell & Poth, 2018). In addition, a focus group can encourage feedback from individuals who may be reluctant to be interviewed in a one-on-one session (Creswell & Poth, 2018). With this approach of open question and answers the researcher had the opportunity to have a firsthand understanding of the defined processes, undefined gaps, tools, and strategy utilized by these stakeholders in obsolescence management. This focus group found success in the ability of participants to build off of one another's commentary and anecdotal examples, to enable better understanding of the lived experience of managing obsolescence issues.

Discussion of the Focus Group

Prior to beginning the analysis, I read and reread the text of the focus group to ensure I understood the participants' responses, patterns, and could get an initial impression of the recurring themes. The coding was conducted through Atlas.ti, an analysis tool. The initial output resulted in 36 quotations from the focus group that were coded to 156 words and phrases around the questions of obsolescence. Of the 156 codes, 16 were identified as the top applied codes, including risk realization, lifetime buys, and obsolescence as the top applied codes with three quotations each.

Results

Upon review of the output, many of the codes were identified to be duplicated themes and were aggregated. The coded questions were revisited to identify potential overlap with other codes and identify trends within the data. The end result was 17 coded concepts with the top four codes as cost impact, securing supply, risk, and collaboration. Figure 1 identifies the 16 themes and aggregated response count for the full focus group session.

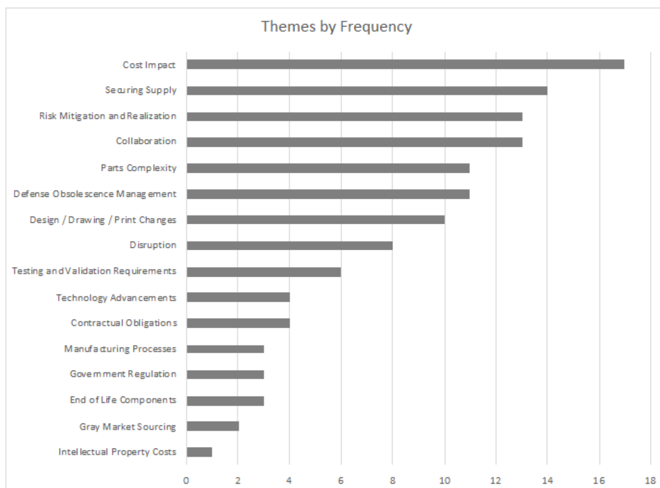


Figure 1. Recurring Themes

The theme of cost impact was identified in the responses of all focus group participants. It was noted that there are different approaches to resolve obsolescence that can be costly and time consuming, including design changes and lifetime buys. One noted comment from one of the participants indicated how obsolescence that requires redesign and involves Intellectual Property (IP) can be pricey. He said:

> We had a situation about a year ago where we actually had a product that had been obsolete for a number of years and the tech IP was acquired by a third party. . . . So, pricing for that material or the alternative to recreate was in the range of about $2,000,000 between the design costs and the material itself to support what was remaining the life of that product. And then we also had the option of the third party that had acquired the IP and was able to go back and produce the original part. In that case. That particular buy I think was roughly $1.5 million. (Respondent x)

In addition to the costs for purchase or design of alternatives, costs can occur from production and factory line down instances while awaiting obsolescence resolution. Figure 2 identifies the top five co-occurrences of concepts with cost impact.

#	Code 1	Code 2	Occurrence
1	Cost Impact	Securing Supply	14
2	Cost Impact	Collaboration	13
3	Cost Impact	Risk Mitigation & Realization	13
4	Cost Impact	Defense Obsolescence Management	11
5	Cost Impact	Part Complexity	11

Figure 2. Cost Impact Co-Occurrences

Next, the importance of securing supply and mitigating risk was noted in multiple responses and coding. One of the focus group study participants described his experience with the complexities of being able to ensure the source of supply and mitigate the risk of obsolescence. He said:

> A lot of times we do find that parts go obsolete not because the supplier can't make it, but because they can't get components or raw materials. I can think of certain instances like in Europe where they're now introducing REACH. A lot of the chemicals used in the manufacturing of a lot of products are no longer going to be available. If suppliers, especially engineered goods suppliers, the ones who own the IP don't get out in front of that, then you know we can have some potential issues in the future. (Respondent x)

Proactive and strategic methods are identified as the best practices to ensure that the impacts of obsolescence can be mitigated. Focus group participants also discussed initiative-taking planning, forecasting, and long-term mitigation actions to reduce the impacts of obsolescence issues and ensure supply continuity. Figure 3 identifies the top five co-occurrences of concepts with risk mitigation and realization.

#	Code 1	Code 2	Occurrence
1	Risk Mitigation & Realization	Cost Impact	17
2	Risk Mitigation & Realization	Securing Supply	14
3	Risk Mitigation & Realization	Collaboration	13
4	Risk Mitigation & Realization	Defense Obsolescence Management	11
5	Risk Mitigation & Realization	Part Complexity	11

Figure 3. Risk Mitigation and Realization Co-Occurrences

Another key theme of this focus group discussion is collaboration. The coding sheet identifies 13 instances of collaboration that includes the importance of coordination between departments and functions internal and external to their organization. Through this approach, expertise and resources can be shared and leveraged to develop a comprehensive approach for obsolescence management. Figure 4 identifies the top five co-occurrences of concepts with collaboration.

#	Code 1	Code 2	Occurrence
1	Collaboration	Cost Impact	8
2	Collaboration	Risk Mitigation & Realization	6
3	Collaboration	Disruption	4
4	Collaboration	Contractual Obligations	3
5	Collaboration	Part Complexity	3

Figure 4. Collaboration Co-Occurrences

In addition, as identified in Figure 4, collaboration must occur in issues of obsolescence to reduce cost impact, risk, and disruption. Also, issues involving contractual obligations and part complexity can result in the need for more collaborative working engagements to align across functions and resolve the obsolescence problem.

The focus group session provided additional insight to supplement the research in the area of obsolescence and factors of cost. Through this approach of open question and answers the researcher had the opportunity to have a firsthand understanding of the defined processes, undefined gaps, tools, and strategy utilized by these stakeholders in obsolescence management and cost impact.

The results of the focus group and analysis have helped to identify further areas of research and exploration for this dissertation. This session provided a valuable bridge between the background archival information of the literature review and the planned study through interview. The focus group participants noted the significant cost impact associated with obsolescence, which include expenses related to design changes, lifetime buys, and product redesigns. Additionally, cost challenges related to material sourcing and contractual requirements were also identified. Proactive planning and long-term strategy were identified as best practices to minimize impact. In addition, the focus group provided a unique opportunity to tap into the knowledge and insight of experienced individuals that has helped guide the researcher to the more focused interview study that will support the study to determine the cost drivers of obsolescence in A&D.

While prior research has explored themes such as proactive management, financial impact, supply chain risk, and strategic forecasting, it has largely overlooked the specific cost drivers contributing to obsolescence impact. The focus group participants who were professionals from various functional roles across A&D consistently emphasized a lack of clarity around the root causes and traceability of costs in obsolescence. Their input affirmed the need to examine not just the occurrence of obsolescence, but the underlying cost factors that result in impact. However, the focus group findings also revealed that while these problems are widely experienced, systematic knowledge and organizational alignment on the issue remain limited, underscoring the need for a more in-depth investigation.

Summary of the Results

The focus group conducted provided valuable insights into the challenges of obsolescence management in A&D. The key themes identified from the participants' discussions centered on cost impact, securing supply, risk mitigation, and collaboration. Participants emphasized the significant costs associated with obsolescence, particularly when redesigning or purchasing obsolete components, and the complexity of securing supply due to limited sources and regulatory changes. The group also highlighted the importance of risk mitigation strategies and the need for stronger collaboration across internal and external stakeholders to effectively manage obsolescence challenges. The focus group served as an important foundation but the topic of obsolescence cost impact should be explored further.

Themes

The focus group data revealed several prominent themes that shape the understanding of obsolescence management in the A&D industry. The dominant theme was cost impact, where participants discussed how obsolescence led to high costs, including redesign efforts, last-time buys, and the procurement of alternative parts. The theme of securing supply was also critical, as participants noted the challenges posed by the shrinking supplier base and regulatory changes, such as the introduction of REACH in Europe. Risk mitigation and impact emerged as a key theme, with participants advocating for proactive strategies to identify and address obsolescence risks before they disrupt operations. Finally, participants identified collaboration as being essential for managing obsolescence, with the need for more coordination between cross functional teams and parties, such as procurement, engineering, and external suppliers, to ensure effective solutions.

Original Contribution to Knowledge

This study's original contribution lies in its qualitative exploration of the cost drivers associated with obsolescence in A&D, a subject that remains under-researched. The focus group highlighted the significant gaps in existing literature, particularly regarding the direct and indirect costs of obsolescence. By identifying the key themes and challenges faced by procurement and supply chain practitioners, this research contributes a nuanced understanding of the costs involved in managing obsolescence, beyond what is typically addressed in traditional frameworks.

The findings from the focus group have several practical implications for professionals working in procurement and supply chain management within the A&D industry. First, they underscore the need for proactive planning in managing obsolescence risks. Organizations are encouraged to integrate obsolescence management early into their procurement strategies, focusing on long-term sustainment through tools like predictive analytics and technology roadmaps. The importance of cross-functional collaboration also emerged, highlighting the need for coordinated efforts across departments to address obsolescence before it escalates into a major issue. Finally, the study calls for greater visibility and communication with suppliers to mitigate risks associated with component discontinuation and supply chain disruptions.

Limitation of Study

While this study provides valuable insights into obsolescence management, it is not without limitations. The focus group was relatively small, with only seven participants, which may limit the generalizability of the findings. Additionally, the study focused exclusively on procurement and supply chain practitioners in the A&D industry, excluding other functional areas such as engineering, quality, and finance, which may also contribute to obsolescence management. Furthermore, the data collected is qualitative in nature, meaning it is more subjective and may be influenced by individual biases or recall inaccuracies. These limitations highlight the need for further research with a larger, more diverse sample and additional methods of data collection to strengthen the findings.

Future Research Potential

Future research could expand on this study by incorporating quantitative data to supplement the qualitative insights and provide a more comprehensive understanding of the cost impacts of obsolescence. Further studies could also explore the perspectives of other functional areas, such as engineering and finance, to provide a more holistic view of the challenges faced by organizations in managing obsolescence. Additionally, the integration of digital tools such as AI-driven predictive analytics, which were not fully explored in this study, could be examined as a potential avenue for enhancing obsolescence management strategies. The insights from this study could be tested across different industries and geographical contexts to assess the broader applicability of the findings.

Conclusion

In conclusion, the focus group has provided valuable insights into the factors that drive the cost of obsolescence mitigation, management, and risk realization in A&D. The findings emphasize the need for proactive strategies, cross-functional collaboration, and increased supply chain visibility when managing obsolescence.

This study bridges the gap between existing obsolescence frameworks and the real-world experiences of procurement and supply chain practitioners, offering new insights into the cost drivers involved. These findings will inform next phase of the research, which will employ a grounded theory approach to explore these issues further through interviews with a larger sample of A&D professionals. Ultimately, this research aims to provide practical recommendations for reducing the high costs associated with obsolescence in A&D and contribute to the development of more effective management strategies.

APPENDIX A

Participant Details

#	Name	Company	Role	Previous Involvement with Obsolescence
1	B.L.	1	Sr. Mgr Outsource Procurement Group	17 years at company 1; last 4 managing obsolescence group
2	G.B.	1	Supplier Development Manager	12 years in industry; 2 years working with shortages (some obs related)
3	J.M	1,3	Sr. Tech Fellow in Global Strategic Sourcing	41 years at company 1 and 3; many years of obsolescence experience
4	M.P	1	PBL Contracts Support for Obsolescence	25 years with company 1 in ops roles; currently supporting obsolescence
5	S.S	2	Sr. Mgr Supplier Performance	24 years company 2; primarily on defense end w/issues of obsolescence
6	S.A	1	Mgr Obsolescence Group	10 years supporting obsolescence group in company 1
7	K.R	1	Procurement	6-7 years of obsolescence procurement in company 1

APPENDIX B

Focus Group Protocol

1. What is your experience managing or having responsibility to identify and resolve obsolescence part issues in the aerospace and defense industry? Provide details such as how many years you have worked in this area, instances of your involvement, how you were involved, etc.
2. Describe an occurrence of parts obsolescence that you managed. Provide details of how the issue was identified, what impact it had on production and the program it supported, the year it occurred and how long it took to resolve the obsolescence issue. Provide any details of financial impact including loss of sales, cost of goods, contract penalties, lost time, etc. Was the material obsolescence planned or unplanned? Known beforehand?
3. What stage of the product life cycle did the obsolescent part fall under? What commodity does this part fall under? Were any of the components, material, parts, product, or technology on this item proprietary to the USG, foreign government, or contractor Was this a commercially available item? What was determined to be the root cause of this issue?
4. What were the total costs for resolution of the obsolescence issue?
    a. Procurement of new materials
    b. Engineering and Drawing Revisions
    c. Testing and Validation
    d. Qualification
    e. Tooling, Equipment, and Software
    f. Other associated costs?
5. Describe any other factors, thoughts, comments related to your experience with this issue of obsolescence.
6. Do you, your team, or others in product, part, and supply chain coordinate within their business, with their customer, with others in industry on obsolescence planning and

mitigation strategies? (DMSMS, etc.)?

References

Bartels, B., Ermel, U., Pecht, M., & Sandborn, P. (2012). *Strategies to the prediction, mitigation, and management of product obsolescence*. John Wiley & Sons.

Blanchard, B. S. (2004). *Logistics engineering and management* (6th ed.). Pearson.

Chellin, M. D., & Miller, E. E. (2023). Proactive obsolescence management methods for C5ISR systems: Insights from practitioners. *Defense AR Journal, 30*(1), 24–44. https://www.dau.edu/sites/default/files/Migrate/ARJFiles/ARJ103/ARJ103_Chellin.pdf

Chellin, M. D., & Gallegos, E. E. (2024). Visualization tool for obsolescence management. *Defense Acquisition Research Journal, 31*(1), 106–123.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches*. SAGE Publications.

Del Campo, H., Poltronieri, G., Simoncini, A., Vaghi, A., & Vesco, S. (2022). How industrial and aerospace and defense OEMs can win the obsolescence challenge. https://www.mckinsey.com/industries/aerospace-and-defense/our-insights/how-industrial-and-aerospace-and-defense-oems-can-win-the-obsolescence-challenge

DoD (2022). *DoD's microelectronics commons takes shape*. https://www.defense.gov/News/Releases/Release/Article/3210295/dods-microelectronics-commons-takes-shape/

DoD. (2024). *SD-22 diminishing manufacturing sources and materials: A guidebook of best practices for implementing a robust DMSMS management program*. https://quicksearch.dla.mil/qsDocDetails.aspx?ident_number=275490

English, W. (2022). *Strategies to mitigate obsolescence risks for U.S. Army precision guided missiles* [Dissertation, Defense Acquisition University Senior Service College Fellowship]. https://apps.dtic.mil/sti/pdfs/AD1180884.pdf

FAA. (2015). *Obsolescence and life cycle management for avionics*. https://www.faa.gov/sites/faa.gov/files/aircraft/air_cert/design_approvals/air_software/TC-15-33.pdf

Howard, M. A. (2022). *Component obsolescence it's not just for electronics anymore*. ARINC, Inc. https://smapcenter.uah.edu/SMAP-CENTER/Conferences/dmsms02/papers/howard.pdf

Jennings, C., Wu, D., & Terpenny, J. (2016, June 27–July1). *Forecasting obsolescence risk using machine learning* [Conference session]. ASME 2016 International Manufacturing Science and Engineering Conference, Blacksburg, VA, United States.

Ozkan, B., & Bulkan, S. (2016). COTS parts obsolescence management of sustainment dominated military systems. *10th NATO Operations Research and Analysis*, 1–18.

Porter, K., & Plotkin, E. (2013). *Four obsolescence management myths that kill defense programs.* Military Embedded Systems. https://militaryembedded.com/radar-ew/signal-processing/four-myths-kill-defense-programs

Rojo, F. J. (2010). Obsolescence management for long-life contracts: state of the art and future trends. *International Journal of Advanced Manufacturing Technology*, 1235–1250.

Sandborn, P. (2013). Trapped on technology's trailing edge: Managing electronic part obsolescence. *Engineering Management Journal, 25*(4), 16—26. https://doi.org/10.1080/10429247.2013.11431992

Sandborn, P., Prabhakar, & Eriksson, B. (2008). The application of product platform design to the reuse of electronic components subject to long-term supply chain disruptions. *2008 ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, DERC 2008, 1*, 989–998.

Solomon, R., Sandborn, P., & Pecht, M. (2000). Electronic part life cycle concepts and obsolescence forecasting. *IEEE Transactions on Components and Packaging Technologies, 23*(4), 707–717. https://doi.org/10.1109/6144.888871

An Assurance Educated Workforce Is Critical to Addressing Software and Supply Chain Acquisition Lifecycle Risks

Dr. Carol Woody—is a Principal Researcher for the CERT division of the Software Engineering Institute at Carnegie Mellon University. She leads a team that builds capabilities and competencies for measuring, managing, and sustaining software assurance and cybersecurity for highly complex software-intensive systems and supply chains throughout the acquisition lifecycle. She coauthored a book, Cyber Security Engineering: A Practical Approach for Systems and Software Assurance, published by Pearson Education as part of the SEI Series in Software Engineering. The CERT Cybersecurity Engineering and Software Assurance Professional Certificate is based on the research she led. [cwoody@cedrt.org]

Abstract

Today's systems are software-intensive and complex, with a growing reliance on third-party technology. Through reuse, systems can be assembled faster with less development cost. Traditionally, systems were hardware-based, and operational risks were primarily linked to reliability. Now systems are largely software-based, which does not wear out like hardware, and the critical risks are different. All software contains vulnerabilities that are hard enough to manage directly. Inheritance through the supply chain increases the management challenges and magnifies the risk of a potential compromise. Attacks on the software supply chain are increasingly frequent and devastating. Software risk management capabilities are brought in too late, if at all, to identify and address software risks that can appear throughout the lifecycle. Extensive compliance rules have been put in place for federal acquisitions to address software and supply chain risk, but there is a noticeable gap in the current acquisition and engineering workforce's knowledge and skills needed to address the rules effectively. Expanding the knowledge of decision-makers and participants in system acquisition, engineering, and integration are critical activities that are necessary to address the growing software risk.

Introduction

Today's systems are increasingly software-intensive and complex, with a growing reliance on third-party technology. Through reuse, systems can be assembled faster with less development cost. Traditionally, systems were hardware-based, and operational risks were primarily linked to reliability. Now systems are largely software-based, which does not wear out like hardware, and the critical risks are different. All software contains vulnerabilities that are hard enough to manage directly. Inheritance through the supply chain increases the management challenges and magnifies the risk of a potential compromise. In addition, suppliers can become propagators of malware and ransomware through features that provide automatic updates. Attacks on the software supply chain are increasingly frequent and devastating.

Extensive compliance rules have been put in place for federal acquisitions to address software and supply chain risk, but there is a noticeable gap in the current acquisition and engineering workforce's knowledge and skills needed to address the rules effectively. Each program develops their unique risk management processes and practices, many of which ignore software. The right capabilities are brought in too late, if at all, to identify and address software risks that can appear throughout the lifecycle. Acquisition and program management are focused on budgets, cost, and schedule, motivating the adoption of shortcuts even in addressing compliance. Expanding the knowledge of decision-makers and participants in system acquisition and engineering is a critical component in addressing the growing software risk, but it does not appear to be anyone's responsibility.

In his memo from March 6, 2025, on *Directing Modern Software Acquisition to Maximize Lethality*, the Secretary of Defense (2025) noted the following:

The Department of Defense (DoD) has been slow to recognize that software-defined warfare is not a future construct, but the reality we find ourselves operating in today. Software is at the core of every weapon and supporting system we field to remain the strongest, most lethal fighting force in the world. While the commercial industry has rapidly adjusted to a software-defined product reality, DoD has struggled to reframe our acquisition process from a hardware-centric to a software-centric approach.

Unfortunately, recent experience at Carnegie Mellon University Software Engineering Institute (CMU SEI) for weapon and support systems has shown that directing the resources that have efficiently handled the DoD acquisition process for decades to now focus on software will not be sufficient. Software is designed, built, integrated, managed, and supported differently from hardware. Current processes and practices have not been tailored and integrated for effectively addressing these differences. Software is also not isolated to specific segments of the system but has become a major portion of virtually all aspects of system development and delivery, requiring an integrated perspective for effective management.

The various participants in an acquisition program focus on their existing areas with expertise limited to their current functions and are only tied to other areas by processes that share data, documents, or dollars to efficiently deliver capabilities as they continue functioning under a hardware-oriented approach. Figure 1 provides a high-level view into the structure of a major acquisition.

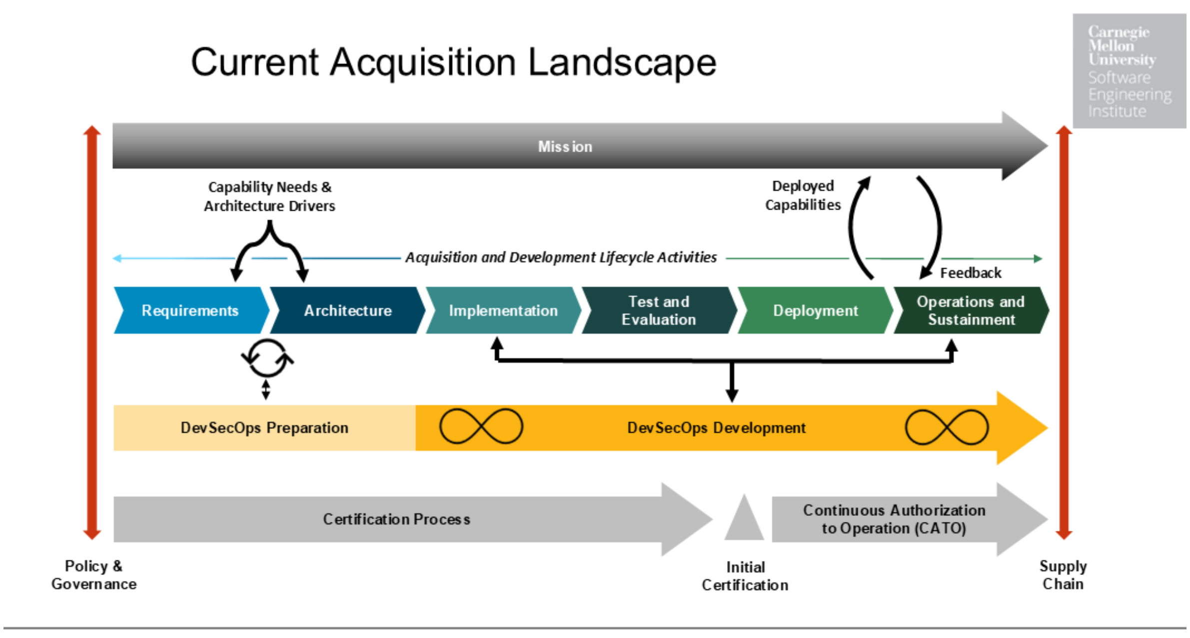


Figure 1. Current Acquisition Landscape

The program management for a major DoD acquisition focuses on the mission and the warfighter capabilities needed to expand mission capabilities (top line in Figure 1). Their primary acquisition product is the statement of work (SOW), which lays out the range of requirements that a contractor will be asked to deliver (including mandatory policy guidance drawn from the line up the left of Figure 1 and compliance drawn from the bottom line in Figure 1). From the SOW, system engineers working for the government (addressing the acquisition and development lifecycle activities in Figure 1) will develop a system specification and system engineering plan (SEP) as part of the activities in the second layer of the acquisition to decompose the SOW with further detail about the engineering rigor expected from the vendor. In parallel, cybersecurity experts (bottom line on Figure 1) are creating a program protection plan (PPP), which is expected to contain a cybersecurity strategy (CSS) defining the requirements for security controls on the system. Engineers from the contractor will use the SOW, SEP, and PPP to build a software development plan (SDP). From the SDP, the contractor—and government, if the final product will be owned by the government—will assemble the tools in development pipelines to automate, as feasible, the software production process (DevSecOps line in Figure 1). There may be many development pipelines addressing various classification levels of the system and software. Software is frequently missing from early milestones, and feedback loops from cybersecurity monitoring and software development to program cost and schedule typically do not exist. These challenges can lead to risk and cost impacts that continue into operations where they are much more costly to address.

At various milestone reviews sprinkled within the acquisition schedule, participants evaluate progress through various processes that involve the review of documents delivered by the contractor. These interactions involve many steps, multiple organizational entities, suppliers, and document exchanges. Eventually they can result in a poorly managed project plan with uncertain cost and schedule milestones. Lacking the knowledge and required integration across government teams necessary to effectively deliver a system in the software-intensive environments we face today, the system can be plagued by inconsistent processes, disjointed compliance-driven risk management plans, and cost-schedule overruns. A more troubling outcome can be a variety of separate views that do not ensure agreement as to what is delivered.

DoD Instruction 5000.02 (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2020) establishes the management framework for translating mission needs and technology opportunities, based on approved mission needs and requirements, into stable, affordable, and well-managed acquisition programs that include weapon systems and automated information systems (AISs):

> To achieve those objectives, Milestone Decision Authorities (MDAs), other Decision Authorities (DAs), and Program Managers (PMs) have broad authority to plan and manage their programs consistent with sound business practice. The AAF acquisition pathways provide opportunities for MDAs/DAs and PMs to develop acquisition strategies and employ acquisition processes that match the characteristics of the capability being acquired. (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2020, p. 4)

Both software and its assurance are latecomers to the acquisition lifecycle. As shown in Figure 1, software’s role has been assigned primarily to the bottom two layers of the structure: the pipeline where it is built and the certification process. Unfortunately, many decisions that impact software are made early in the acquisition by system engineers, contracting, and supply chain management that directly impact the assurance of the software, but those with software expertise are typically not included in these earlier steps.

Systems engineers have dominated the early stages of the acquisition and development lifecycle in both the government and defense contractor organizations following the Department of Defense Architecture Framework (DoDAF) principles (DoD Chief Information Officer, 2010), creating elaborate overview and detail diagrams that show how a new system will be interfaced with existing capabilities and built from components. Software has long been relegated to the lower tiers of the DoDAF. Systems engineers decompose the capabilities into independent components, following good engineering practice (INCOSE, 2023) to reduce the complexity needed to be considered as each individual component is built. The assembly of the components is expected to yield the desired capabilities with the desired qualities (such as security and safety), which are emergent properties of the integrated whole. Each component is assigned functional requirements that flow down from the SOW and SEP. Interfaces among components are assumed to be well formed and isolated. Risks are evaluated at the system level, and appropriate controls are incorporated into the design to ensure the requirements for security based on confidentiality, integrity, and availability are met, and these will be evaluated for compliance by cybersecurity experts before the system will receive an authority to operate (bottom line in Figure 1). Too frequently, the processes are not well managed and integrated by individuals with the skills necessary to ensure that software considerations are addressed appropriately.

As part of the modernization planning, the DoD is migrating storage usage to cloud services and outsourcing other capabilities to reduce infrastructure costs and enhance enterprise capabilities to share information (DoD Chief Information Officer, 2024). These services are primarily software-based, further increasing the layers of software incorporated into an acquisition. These choices are typically made by program management and shift the control of this software from system engineering to supply chain management, but software expertise is not typically part of this team. The third-party software components may include additional capabilities that expand the available functionality and external interfaces and violate the independence of the components.

The DoD is also embracing modernization of software development using Agile techniques for incremental development and software factories. Tools for the factory pipelines may augment the code such that the functionality delivered goes beyond the original requirements, which can violate isolation assumptions and result in the inclusion of capabilities that allow bypassing of security controls implemented at the system level. Unless the system engineers prepare the system design to be implemented incrementally, the software factory selection of what will be done and in what order will be made at the software development stages. Choices for the sequence of what portions are developed may determine the readiness of the system for meeting the system's qualities (such as security and safety) and conflict with program management expectations for implementation.

When software is created, available modules and code libraries from third parties that provide the functionality needed are extensively reused, creating an unexpected dependency on the supply chain. Management and oversight of those suppliers is frequently overlooked due to lack of software expertise and skills. Software components are often interrelated sets of functionality (one layer is *not* necessarily contained inside another layer), and routines that address shared functionality are created as shared subroutines interfaced to multiple modules instead of repeated inside each of the use points. This minimizes the maintenance requirements of the code in the future since all uses are taken from one source but violates the independent assumptions of each individual software component inherent in the system design.

Gaps in Program Knowledge About Software Risks

As noted earlier, software is designed, developed, managed, and monitored uniquely. Software is intellectual property and is the output of creativity and knowledge of its writer. A reality with software is that all software contains potential vulnerabilities: either inserted through gaps in the language structures if secure coding standards are not enforced or inherited from reused components—or both. For many third-party software products and open-source products, these vulnerabilities are publicly available through the National Vulnerability Database (National Institute of Standards and Technology [NIST], n.d.). Too frequently, there is a lack of recognition of the risks these vulnerabilities represent to the program. The PPP should include considerations for software risks, but software expertise is typically missing in the supply chain risk management teams that have responsibility for this document. Cybersecurity may be enforcing the NIST Risk Management Framework (RMF; NIST, 2018), which includes recommendations for vulnerability identification through the application of static and dynamic analysis tools. However, to reduce the vulnerability risk to a program, the requirements for the acquisition must include removal of these vulnerabilities; this removal must be incorporated into either the pipeline activities for the software factory or managed through the input to the pipeline as a backlog entry. Handling of vulnerabilities in third-party software, which a program does not directly control, may require software design constraints. These constraints need to be managed in a program software architecture, which is too frequently disbursed into each software component as part of the system architecture without consideration of system-wide needs that should be integrated across the program.

Software products cannot be implemented and ignored. Few programs recognize the realities of software reliability that must be constantly monitored for obsolescence, changes in business needs that require adjustments, and new vulnerabilities discovered by others and published, increasing the risk to those still using the software. Even if the risks are identified and reported, risk management procedures are too frequently not integrated with software management activities. At the SEI, we see many organizations in which software risks are reported and collected when software is being developed, but the organizations lack mechanisms for escalating these risks to program decision-makers. Lacking an awareness of the software risks, program leaders do not know when and how to respond until a crisis occurs.

Program management monitors the cost and schedule for the acquisition and is focused on effective delivery of the requirements as defined in the SOW. Too frequently, the development of the SOW does not integrate cybersecurity, software assurance, and software supply chain risk management requirements. Even when these requirements are included, personnel knowledgeable in software and cybersecurity are typically not part of the early lifecycle activities; therefore, consideration in the early planning and engineering is missed. Too frequently, the SOW will require meeting such specific policies as AFMAN 91-119 (DefenseMirror.com, 2024) and NIST 800-53 (NIST, 2018) and require the contractor to address the RMF as the cybersecurity requirements. These policies and standards are written at a general guidance level that must be tailored to specific risk considerations for the program; however, without the proper expertise, appropriate tailoring is not happening. The contractor may select controls that are insufficient for the actual risks without providing clarity as to the specific software and cybersecurity concerns to be addressed, driven by a compliance mentality, without adequate tailoring to address the software-related risks. In other instances, only external system risks that are mandated for compliance are considered, and software risks that are based on supply chain decisions (made by both the prime contractor in handling their subcontractors and the government in their software supply chain) are overlooked.

Program Needs for Risk Management of Software

Having the right knowledge to recognize and understand cybersecurity and software risks throughout the acquisition and development lifecycle is critical. Program management, systems engineers, and supply chain acquisition resources need to understand the risks to the program/system and appropriately identify and manage them throughout the lifecycle. This knowledge is not currently part of the expertise required for these positions. Having an effective risk management framework in place to connect software risks with the handling of program risks is critical to the success for programs with intensive software and software supply chain components.

Program leadership and acquisition personnel need to know how to address the following issues: (1) When do we need to include software, cybersecurity, and software supply chain expertise? (2) How do we get the resources we need at the right place in the program to address the growing needs for software and supply chain risk management (SCRM) with a workforce that is currently not prepared to handle these responsibilities and a pipeline of future workers who have never heard about software vulnerabilities in their education, much less learned how to address them? In addition, program leadership must understand that responsibility for software is widely scattered across all parts of the acquisition and development lifecycle, and collaboration among these various players is typically nonexistent. The ability to build and manage the processes that are required for software-intensive systems is essential and requires that software informed expertise and training become a priority.

Different program groups develop the SOW, SEP, and PPP. When software and cybersecurity are included, they need to be consistent and integrated, and in most cases, we have seen wide discrepancies among the requirements in each of these documents. At a minimum, we must raise the awareness of leadership that software-intensive systems require new skills, training, and an expanded management mindset. Today's acquisitions are increasingly software-intensive, complex, and reliant on third-party technology (i.e., hardware, software, and firmware).

The strategic transition to commercial software can serve to expand software risk management to a more lifecycle-oriented perspective. Programs will rely more on vendors that address security issues through patches and upgrades that must be constantly monitored and integrated. As vendors release new versions, older products are no longer maintained, and existing vulnerabilities are not addressed. Programs are not currently structured to continuously update third-party and open source software products. Obsolescence will be a growing issue for an environment that is accustomed to long implementation cycles. The DoD leadership guidelines do not come with consideration of the shift in responsibilities to address the expanded role that software-intensive systems bring. Risk decisions made by acquisition personnel must expand beyond the lowest cost and include strategies that address the increased risk posed in a software-intensive environment.

Framework for Effective Software Risk Management

Personnel to address software assurance need to be integrated into every acquisition from the start. These individuals need to understand how systems can be compromised by software; they also need to be aware of mechanisms available for software risk mitigation and how to connect the opportunities for effective management of software concerns into the range of acquisition activities underway at the program and system level

The responsibility for software assurance is laid out by the DoD as follows (DoD Chief Information Officer, 2024):

> *Software Assurance:* The level of confidence that software functions as intended and is free of vulnerabilities, either intentionally or unintentionally designed or inserted as part of the software throughout the lifecycle.

In establishing confidence that the system will be delivered with appropriate software assurance, those addressing software must assemble information from the contractor, government oversight, and across the program early in the lifecycle to predict the level of software assurance that is required based on available evidence and course correct as needed throughout the lifecycle. Later in the lifecycle, software assurance personnel will need to collect data to confirm that results are as expected; this validation will be done in preparation for final verification prior to planned deployment and transition into sustainment activities for monitoring and management of software risk.

SEI researchers, led by the author, have been working with major federal programs to identify effective processes and practices for software assurance and supply chain risk management and have published them in the Acquisition Security Framework (ASF; Alberts et al., 2022). In addition, we conducted two panels at the Software and Supply Chain Assurance Forum: the first in January 2024 on "Establishing the Demand Signal for Good Software Assurance" and the second in May 2024 on "Positioning for Software Assurance Success: Practices, Tools & Technology, Knowledge & Skills." The programs supporting these panel discussions have experience in addressing the challenges of software assurance and software supply chain risk. All of them identify education of program leadership and acquisition integration as primary considerations for success.

It is critical to ensure that the expertise needed is in the right place, and the understanding of the criticality of having this expertise falls on program management. Programs can acquire these capabilities or grow them. In addition, the DoD should consider how to more effectively provide this level of expertise for program use. The Defense Acquisition University (n.d.) is assembling training to support this critical need, but current expertise is limited. There are challenging questions to be addressed by each program:

- Who do we hire or educate?
- What do they need to know to address software and supply chain risk for a program's areas of responsibility?
- How should they learn about what they need to know?
- What expanded collaborations are needed within the lifecycle for the program to provide effective operational results?
- Who is available to the program leadership currently showing success in handling software and supply chain risk to share lessons learned?

Future Considerations

Software vulnerability risk and software supply chain risk are major attack vectors for all technology, and the growth rate is exponential. However, DoD programs appear to not recognize this sufficiently early in the acquisition lifecycle to plan for cost-effective mitigations; instead, consideration is deferred into system integration stages when correcting the gaps is very costly. Because software risk is not well understood by the programs as a key responsibility, DoD funding for specified actions to address software issues is driving the level of consideration provided. As an example, recent mandates to create software bill of materials (SBOMs; Executive Office of the President of the United States, 2022) support improving software supply chain visibility, but programs are claiming this is an unfunded mandate that they are not funded to address. The gaps in understanding software risk and the imperatives for cost-effective execution require an assurance-educated leadership to provide appropriate guidance and an assurance-educated workforce to know how to effectively address the challenges.

Legal Markings

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.

References

Alberts, C., Bandor, M., Wallen, C., & Woody, C. (2022, October 31). Acquisition security framework (ASF): Managing systems cybersecurity risk. *Carnegie Mellon University Software Engineering Institute*. https://doi.org/10.1184/R1/21357627

Defense Acquisition University. (n.d.). *Software acquisition - software acquisition (SWA) resources.* Retrieved April 25, 2025, from https://www.dau.edu/aaf/swa/resources

Defensemirror.com bureau. (2024, May 10). *DoD Inspector General's office evaluates U.S.A.F.'s nuclear design certification for aircraft carrying B61-12 nuclear bomb.* https://www.defensemirror.com/news/36762

DoD Chief Information Officer. (2010, August). *The DoDAF architecture framework version 2.02*. https://dodcio.defense.gov/Library/DoD-Architecture-Framework/

DoD Chief Information Officer. (2024, September 24). *Memorandum for senior pentagon leadership defense agency and DOD field activity directors: Department of Defense cloud financial operations strategy*. https://dodcio.defense.gov/Portals/0/Documents/Library/DoDCloudFinOpsStrategy.pdf

Executive Office of the President of the United States. (2022, September 14). *Memorandum for the heads of executive departments and agencies: Enhancing the security of the software supply chain through secure software development practices* (Report No. OMB M-22-18). https://www.whitehouse.gov/wp-content/uploads/2022/09/M-22-18.pdf

INCOSE. (2023). *Systems engineering handbook: A guide for system life cycle processes and activities, version 5.0*. John Wiley and Sons, Inc.

National Institute of Standards and Technology. (n.d.). *National vulnerability database (NVD)* (SP 800-52 Rev 2). Department of Commerce. https://nvd.nist.gov/

National Institute of Standards and Technology. (2018, December). *Risk management framework for information systems and organizations: A system life cycle approach for* *security and privacy* (SP 800-37 Rev. 2). Department of Commerce. https://csrc.nist.gov/pubs/sp/800/37/r2/final

Office of the Under Secretary of Defense for Acquisition and Sustainment. (2020). *Operation of the adaptive acquisition framework* (DoDI 5000.02). DoD. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500002p.PDF

Office of the Under Secretary of Defense for Research and Engineering, & Office of the DoD Chief Information Officer. (2024, February 16). *Protection of mission critical functions to trusted systems and networks*. (DoDI 5200.44). DoD. https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodi/520044p.pdf

Secretary of Defense. (2025, March 6). *Memorandum for senior pentagon leadership commanders of combatant commands defense agency and DOD field activity directors: Directing modern software acquisition to maximize lethality*. DoD. https://media.defense.gov/2025/Mar/07/2003662943/-1/-1/1/DIRECTING-MODERN-SOFTWARE-ACQUISITION-TO-MAXIMIZE-LETHALITY.PDF

THIS PAGE INTENTIONALLY LEFT BLANK

Conceptual 3D Modeling and Simulation for Wargaming

Shawn Masters—is the Chief Technology Officer for AELIUS. Masters is a technical executive with 30 years of experience as an Electrical Engineer and Computer Scientist in the government and commercial sector pushing technology boundaries to solve unique problems around high performance computing, data processing at scale, and embedded solutions. Masters has worked across industry and the government contracting landscape throughout his career supporting the intelligence community and the Department of Defense in solving perplexing problems. Shawn holds a BS in Electrical Engineering from George Mason University. [shawn@aeliustech.com]

Tim Liberatore—is Senior VP of Engineering and Contracts for AELIUS. A retired Navy Captain and Diver, he served as Commanding Officer of NAVFAC Marianas, Deputy Fleet Civil Engineer at U.S. Pacific Fleet, Ocean Facilities Program Director, Executive Officer at NAVFAC Southwest and the Engineering and Expeditionary Warfare Center, Senior Project Manager at the Research and Development Support Activity, and Commanding Officer Underwater Construction Team TWO. Liberatore deployed for Operation Iraqi Freedom as an Advisor with the Ministry of Interior. He holds a Master of Engineering from U.C. Berkeley, and a BS in Aerospace Engineering from the U.S. Naval Academy. [tim@aeliustech.com]

Abstract

Aelius Exploitation Technologies, LLC (AELIUS) is currently supporting efforts at the forefront of three-dimensional (3D) conceptual modeling and simulation for wargaming in order to advance and transform digital technology to the warfighter.

AELIUS has the demonstrated skill set, resources, and capabilities necessary to develop conceptual models for wargaming in an immersive (augmented reality/virtual reality [AR/VR]–capable), multiplayer simulation, 3D photo-realistic digital synthetic environment. AELIUS is currently contracted with the U.S. Air Force to support a large area test and training range to create an interactive, immersive 3D environment for their entire range area, and recently completed a Defense Advanced Research Projects Agency (DARPA) Phase II effort to create a digital twin of the moon and integrate 3D models of lander, rovers, and infrastructure for dynamic simulation.

AELIUS has managed and integrated wargaming technology into a 3D digital synthetic environment for use in Concept of Operations (CONOPs) development and planning, rehearsals and training, simulations and modeling, specifically in support of Special Operations Forces (SOF). AELIUS has also integrated artificial intelligence/machine learning (AI/ML) models into these synthetic environments, understands how to deal with the largest dataset problems in the environment, and knows how to address the biggest capability gaps in the diversity and robustness of the data available to use for training new algorithms. As such, AELIUS can expand on previous success in combining multimodality real-world data and synthetic data from theater-wide “digital twins” for the robust training of AI/ML models in wargaming applications.

Introduction

The rapid advancement of 3D modeling and simulation technologies, along with AI/ML, presents an opportunity to revolutionize wargaming for military training, strategic planning, and operational analysis. This paper outlines the development of a multi-dimensional/3D wargaming modeling and simulation system to enhance decision-making, improve tactical proficiency, and simulate complex battlefield scenarios with unprecedented realism and adaptability (Tribolet, 2024).

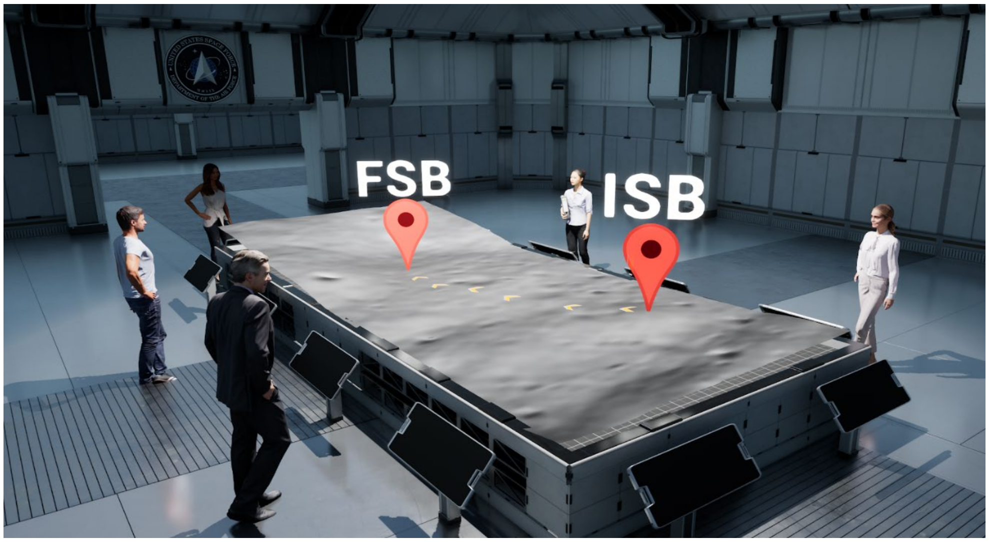


Figure 1. Sand Table in Synthetic Environment for Mission Planning

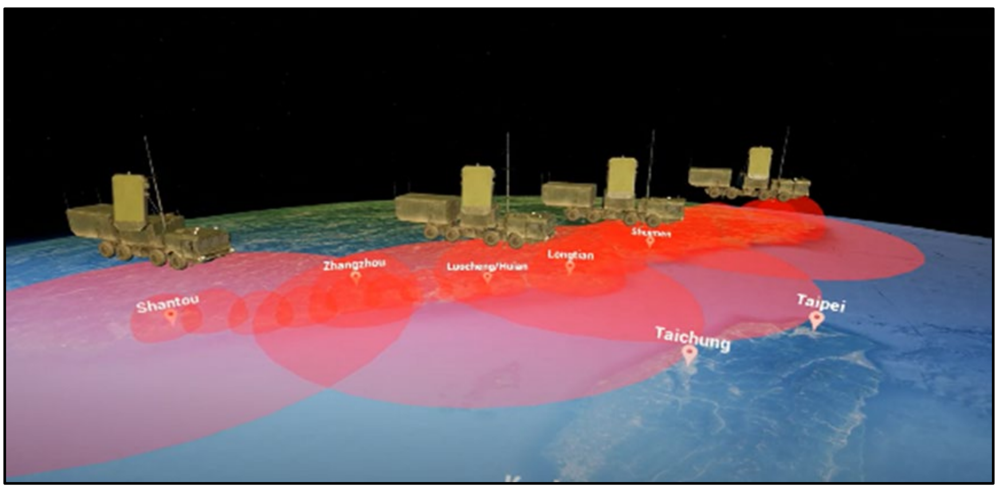


Figure 2. Synthetic Environment of Integrated Air Defense Systems for Mission Planning

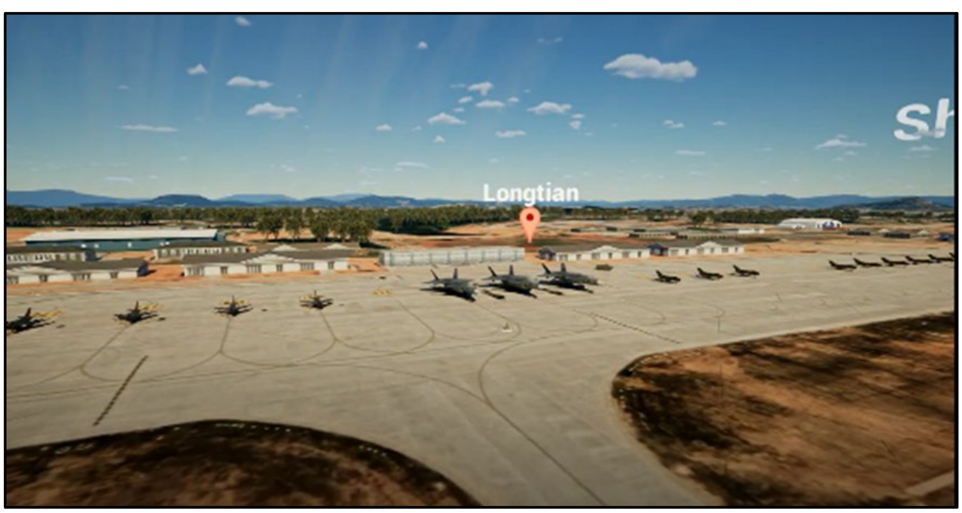


Figure 3. Synthetic Environment of an Airfield

Objectives

- **Photo-Realistic Simulation:** Create a highly detailed 3D synthetic/digital environment that accurately replicates terrain, weather, and unit dynamics for immersive AR/VR wargaming.
- **Scalability:** Design a modular system capable of simulating small-scale tactical skirmishes to large-scale, theater-wide, multi-domain strategic operations.
- **AI/ML Integration:** Incorporate artificial intelligence and machine learning to model enemy behavior, adapt scenarios in real time, and provide post-simulation analysis.
- **Training and Analysis:** Enable military personnel to train in virtual environments and provide commanders with tools to test strategies and predict outcomes.
- **Cost Efficiency:** Reduce reliance on physical resources and live exercises by offering a reusable digital alternative that far exceeds current two-dimensional wargaming constructs.

Figure 4. Synthetic Environment of Taiwan

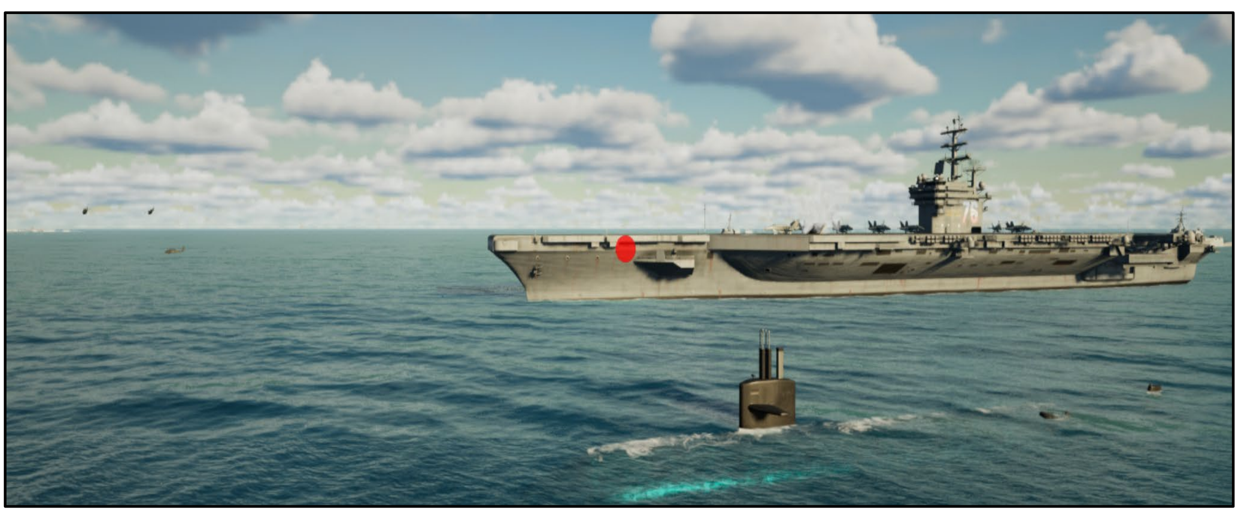


Figure 5. AELIUS Synthetic Environment of U.S. Navy deployed in the Indo–Pacific Theater

Scope

The proposed system-of-systems platform will include

- **3D Terrain Generation:** Procedural and manual tools to create realistic battlefields based on real-world geographic, bathometric, and architectural, data, or fictional landscapes.
- **Unit Modeling:** Detailed representations of infantry, vehicles, aircraft, and naval (surface and undersea) assets with accurate physics and capabilities.
- **Environmental Factors:** Dynamic weather, time of day, and seasonal effects that influence visibility, mobility, and equipment performance.
- **Scenario Editor:** A user-friendly interface for designing custom missions, objectives, and rules of engagement.
- **Multiplayer Functionality:** Support for collaborative and adversarial simulations across distributed networks.
- **After-Action Review (AAR):** Tools to record, replay, and analyze simulations for debriefing and learning.

Figure 6. Integration of AFSIM into UE5 Demonstration

Methodology

AELIUS's solution charts a path for 3D modeling and simulation wargaming, as well as how to integrate and train AI/ML models in contested environments. Even where data collection is possible, but wildly too expensive to gather, there is a wide range of data diversity, and even when the required assets are all owned by the U.S. government, arranging for every configuration option, in every lighting and weather scenario, is not feasible. Even for a simple data set around one vehicle, gathering the data could require weeks to months of work. With some assets and units costing tens of thousands of dollars an hour to operate, there may not be the budget to collect even a small percentage of the required data for accurate modeling, simulation, and training. For a hostile asset, this problem becomes exponentially worse.

Given the limitations of current modeling and simulations, which rely primarily on preestablished scenarios and require extensive human interaction and decision-making, the incorporation of AI/ML offers the capability to simulate numerous scenarios and outcomes. Advances in AI and significant increases in computing power offer opportunities to mitigate the above cost, time, and other identified constraints (Jung, 2024).

With AI-embedded and generative simulation versus traditional simulation, there is the ability to simulate thousands of tactical and strategic battle events in a short period. Through these iterations of play, Jung (2024) proposed AI could "generate vignettes autonomously, produce numerous courses of action for given scenarios, and offer decision-makers multiple options. It also can evaluate or generate optimal actions for opposing forces and devise countermeasures to defeat them."

Figure 7. Simulated Beach Assault Force on Various Islands

A. Research and Requirements Gathering (Phase 1)

   a. Collaborate with military experts to identify key wargaming needs.

   b. Benchmark existing simulation platforms to establish technical standards.

B. Development of Core System (Phase 2)
   a. Use a game engine (e.g., Unreal Engine) for 3D rendering and physics simulation.
   b. Integrate AI frameworks (e.g., reinforcement learning models) for adaptive opponent behavior.
   c. Build initial terrain and unit libraries based on open-source data and licensed assets.

C. Testing and Iteration (Phase 3)
   a. Conduct alpha testing with a small group of military users to refine usability and realism.
   b. To validate system flexibility, simulate a variety of scenarios (e.g., urban combat, distributed maritime operations, and desert operations).

D. Deployment and Training (Phase 4)
   a. Deploy the system to military bases or cloud-based servers for accessibility.
   b. Provide training workshops for operators and scenario designers.

Technical Requirements

- **Hardware:** High-performance GPUs for rendering and AI processing, and VR compatibility for immersive training.
- **Software:** Game engine (Unreal Engine) with robust physics and networking capabilities, AI libraries, and secure data management tools.
- **Data Sources:** Satellite mapping imagery, military equipment specifications, and historical battle records for authenticity.

Technical Summary

Wargaming in a 3D Synthetic Environment Using Unreal Engine

AELIUS's solution develops a wargaming construct within a 3D physics-based gaming environment (Epic Games, Unreal Engine 5 [UE5]) with digital twin models of assets (including equipment, buildings/facilities, aircraft, ships/submarines, vehicles, etc.) and "avatar" personnel, allowing for VR user interface. The digital synthetic environment creates, simulates, and models a photo-realistic presentation of the actual projected combat locations with environmental/weather effects, time of day, and other plug-in features, as required.

AELIUS can continue to enhance and enable USN's wargaming models through visualization enhancement and integration of gaming software (e.g., UE5), as well as AI/ML tools. AELIUS provides the USN a unique blend of technical/scientific and SOF subject matter experts (SMEs) who are practiced in accomplishing operational missions in demanding and austere environments with an in-depth understanding of mission requirements and constraints.

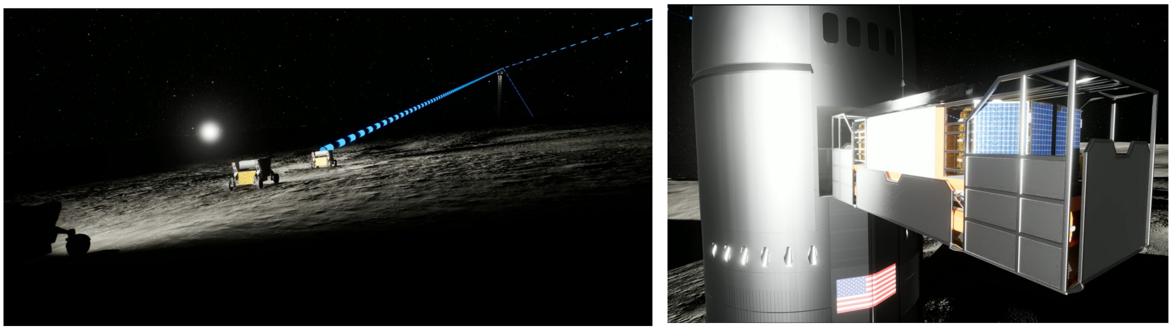
Figure 8. Aelius's Extensive Library Repository of U.S. and Foreign Assets

AELIUS has a resident staff of commercial engineers, scientists/technologists, communications experts, and former SOF operators who can immediately begin supporting 3D wargaming tasks and requirements, as well as a strong bench of additional SMEs and commercial like-minded business partners to meet and exceed Department of Defense (DoD) client-specific data analysis and visualization needs.

Recent Example of AELIUS's Capability of 3D Visualization, Simulation, and Modeling

AELIUS recently created one of the largest landscape data sets for DARPA, comprising over 14.6 million square miles of a photo-real, fully immersive 3D mapping model of the moon's surface. The model integrated NASA's existing maps to simulate future landing sites and lunar surface unmanned operations, with a physics engine replicating the one-sixth scale gravity and incorporating true orbital distances and dynamics for the sun, the moon, and the earth.

Figure 9. AELIUS Synthetic Simulation of Lunar Ice Mine Operations



Figures 10–11. AELIUS Synthetic Environment of the Digital Synthetic Moon Landing Site

Benefits to the DoD

- **Enhanced Preparedness:** Train personnel in diverse, repeatable scenarios without logistical constraints.
- **Strategic Insight:** Test hypotheses and refine tactics in a risk-free environment.
- **Cost Savings:** Minimize expenses associated with live exercises, such as fuel, ammunition, and equipment wear.
- **Adaptability:** Quickly update simulations to reflect emerging threats or new technologies.

Conclusion

AELIUS's proposed 3D wargaming modeling and simulation platform offers a transformative tool for military training and strategic analysis. By leveraging cutting-edge technology, this system-of-systems will empower decision-makers and warfighters to prepare for modern conflicts with greater precision and confidence.

References

Jung, H. (2024, June). A glimpse into the future battlefield with AI-embedded wargames. *U.S. Naval Institute Proceedings*, *150*(6). https://www.usni.org/magazines/proceedings/2024/june/glimpse-future-battlefield-ai-embedded-wargames

Tribolet, J. (2024, January 9). *Wargaming the future: Educating the fleet in multi-dimensional warfare*. Center for International Maritime Security. https://cimsec.org/wargaming-the-future-educating-the-fleet-in-multi-dimensional-warfare/

Building USNA Campus Resilience to Sea-Level Change Effects using X3D Model Publication and Visualization in SPIDERS3D

Don Brutzman, PhD—is a computer scientist and Associate Professor of Applied Science working in the Modeling Virtual Environments Simulation (MOVES) Institute, Undersea Warfare Academic Group, and Information Sciences Department. He leads the Network Optional Warfare (NOW) project exploring fleet stealth using efficient messaging, optical signaling, semantic coherence and ethical control of unmanned systems. His research interests include underwater robotics, distributed interactive simulation, real-time visualization using the Extensible 3D (X3D) Graphics International Standard, Artificial Intelligence (AI) and networking for distributed large-scale virtual environments (LSVEs). He is a retired submarine officer who graduated from USNA in 1978. [brutzman@nps.edu]

Abstract

Significant modeling work visualized the United States Naval Academy (USNA) campus by postprocessing 3D data scans using X3D modeling techniques for Web publishing. Evaluating visualizations of projected sea-level rise levels due to climate change first led to improved facilities planning for installation resiliency, and then a decision to raise the campus seawall. Potential cost and capability savings are immense. As an important repeatable exemplar, this planning and visualization project offers significant potential value to many Naval facilities. However, although locally viewed visualization analysis provided much insight, permission to release government-owned 3D data and models was never granted. Upon ensuring that all physical-security concerns are addressed, this paper recommends releasing X3D models and relevant data assets for collaborative visualization by USNA, Naval Postgraduate School (NPS), and the Navy. Careful planning of model creation and model sharing offers an opportunity to broadly support the Navy's climate-action planning activities. Additionally, these assets can lead to ongoing applied education and research work by faculty and students at NPS and USNA. This case study provides multiple lessons learned and recommendations for archivable 3D models with broad implications for the defense acquisition process. This paper describes ongoing efforts, continuing to establish a basis for repeatable 3D visualization as a fundamentally important Naval capability.

Motivation

Climate change is a major destabilizing force that affects operational readiness of the U.S. Navy and Marine Corps. Perhaps the most visible aspect of such changes is sea-level rise, which can have a major impact on port and harbor installations. Critical infrastructure is rightly recognized as it "generates, sustains, and postures the force for the fight." Significant imperatives are provided in references Del Toro (2020) and Francetti (2024). Publication of appropriately curated models and documentation of this methodology can further propel Navy resiliency efforts to mitigate future potential climate-change impacts. This emerging, repeatable capability provides significant power for Navy and United States Marine Corps (USMC) installation resilience.

Project Background and Already-Published Products

The USNA developed an Integrated Adaptation Framework to consider potential climate-change impacts from sea-level rise, as shown in Figure 1. Multiple efforts by Naval Support Activity Annapolis (NSAA) led to contracted work for robot scanning of campus building exteriors, adjacent buildings, and utility tunnels to create an integrated set of models, aggregated for Web-compatible publishing using the royalty-free Extensible 3D (X3D) International Standard.

Using these important assets, simple visualization techniques were then added to show expected impact of sea-level rise, as illustrated in Figure 2 and Figure 3. USNA (n.d., 2022, 2023)[1] describe the rationale for this institutional decision with major scope and impact. Understanding the data-collection products, analytic capabilities, and decision process that was undertaken offers significant potential value for other Naval installations as well.

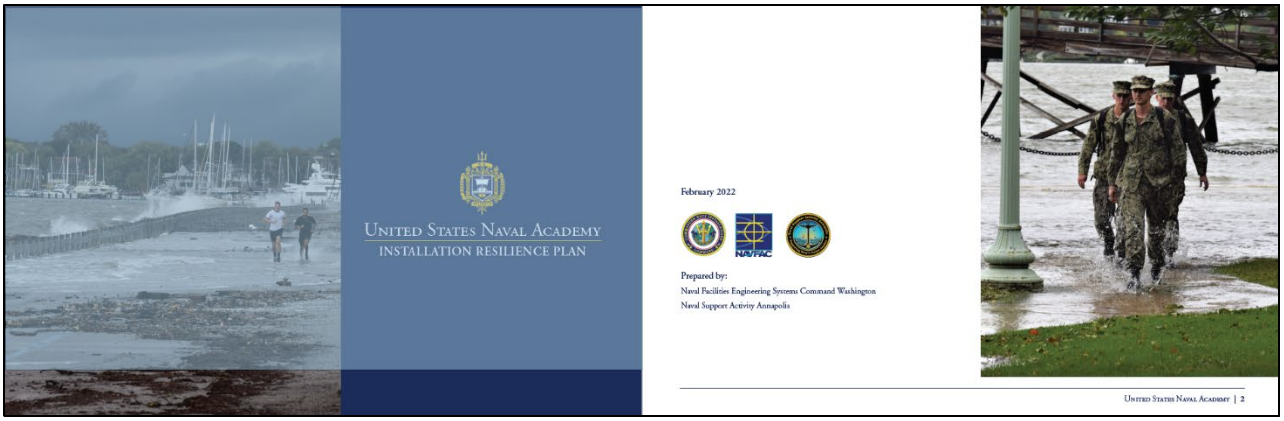


Figure 1. Cover Pages of Publicly Released USNA Installation Resilience Plan 2022 (USNA, 2023)

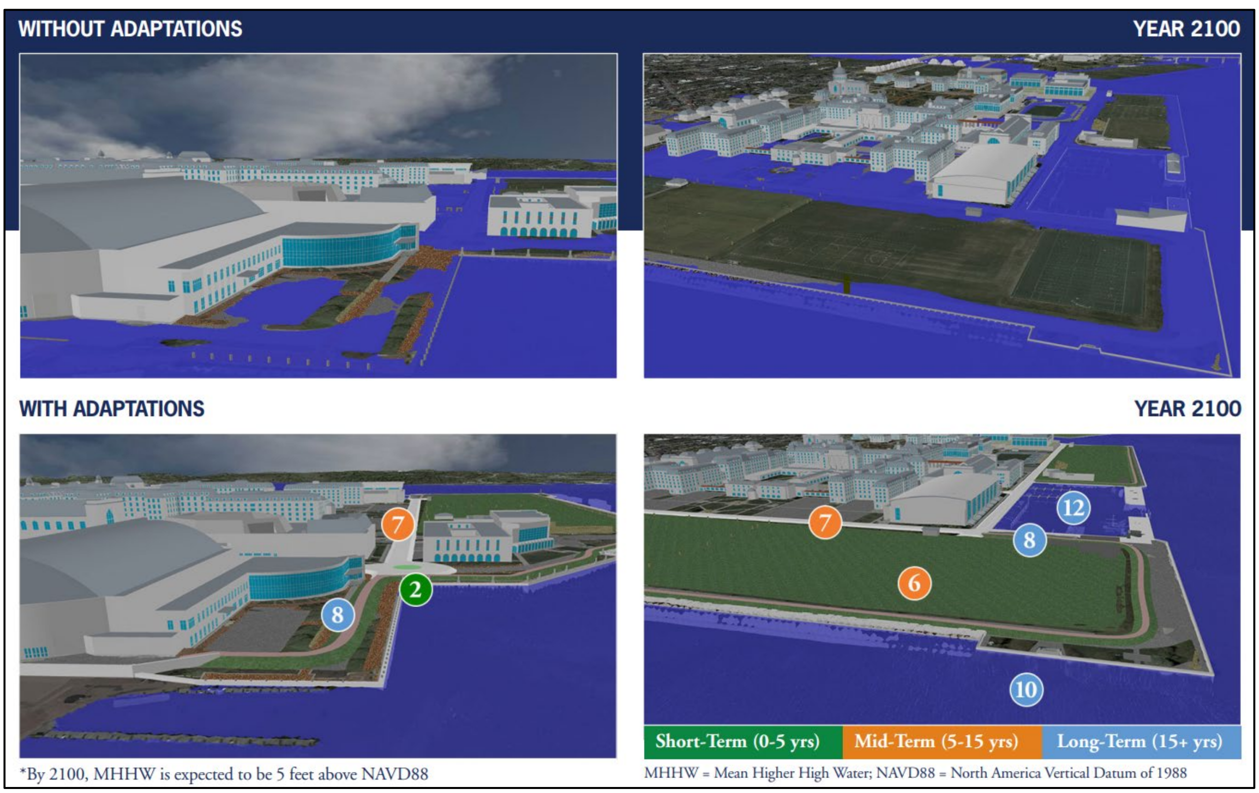


Figure 2. 3D Modeling Portrayals from USNA Installation Resiliency Plan 2022 (USNA, 2023)

[1] The resilience plan, which includes an integrated adaptation framework, project portfolio, and phased execution plan, as well as an executive summary, can be found on the Naval Academy website.

Figure 3. Projected Future Flooding-Level Comparison Using X3D/HTML Modeling, Captured from Demonstration Video

Potential Security Concerns

A number of sensible security concerns have been raised regarding what risks might be associated with the publication of 3D models. Known issues follow, along with corresponding assessments and mitigations for each. Identification of further potential issues is welcome. Following a due-diligence review process during preparation for publication is appropriate for USNA and NPS, each of which hold authorities for handling CUI information. Documenting this process can benefit future ascertainments by other Department of Defense (DoD) installations as well.

Avoiding Unintended Exposure of Sensitive Information

This 3D modeling effort is not cross-connecting databases or exposing any live, critical assets relating to physical security of base facilities. As with publication of any other information asset, content is separated from original sources and then edited or filtered prior to any release.

Avoiding Exposure of Sensitive 3D Models

No campus assets are exposed beyond what is already visible to the many visitors traversing Naval Academy grounds each week. For example, there is no need to publish cabling plans or utility tunnels which conceivably might be exploited by a bad actor. The level of fidelity associated with modeling campus buildings is limited to building exteriors.

Allowing Inclusion of Geospatial Locations

Since any public map can quickly reveal the latitude/longitude location of all facility assets, appropriate internal use of precise geospatial coordinates does not reveal any private or sensitive information.

Allowing Inclusion of Metadata and Appropriate Hyperlinks

The same precautions regarding undue exposure that are used for 3D models need to be similarly applied to accompanying metadata information about those 3D models. Addition of hyperlinks permits users to click on 3D models (buildings, monuments, etc.) to launch an explanatory Web page. Once again, the basic principles of Web design pertain.

Many Related Examples are Available

World Wide Web standards have simplified the creation of hybrid, interlinked, multimedia information sources. The Extensible 3D (X3D) Graphics International Standards enables 3D models to be similarly integrated with other media, effectively making interactive 3D a "first-class capability" on the Web. Of note is that literally hundreds of similar naval facility models are already published via CAC-controlled access to the NAVFAC/EXWC SPIDERS3D system. Similarly the NPS Savage and SavageDefense Examples Archives provide unrestricted or CAC-controlled access, respectively, to numerous additional multimedia 3D models.

Like It or Not: Avoiding a False Sense of Heightened Security

Although prose text, 2D imagery, audio narration, movies, and 3D models provide different presentation modalities and use different formats, they all convey information. If verified information is already available or possible publicly, then hiding government information adds no privacy. Similarly, deliberately remaining unaware of commonly available capabilities does not provide actual security. Security through obscurity should not be used as the only security feature of a system, as explained in "Security through Obscurity" (n.d.).

Deliberate Control of Authoritative Official Confirmation

For a high-profile site like the USNA campus, numerous information assets from varied sources are available publicly. These are sometimes derived from federal, state, public or commercial resources. Recent advancements in technology even enable synthetical AI-produced text, imagery, audio, or 3D graphics. No inappropriate disclosure is expected, based on experienced handling of these matters in accordance with Navy policy.

Technology Basis: X3D and Open Web Standards

Extensible 3D (X3D) Graphics is the open, royalty-free international standard for publishing, viewing, printing, and archiving interactive 3D models on the Web. X3D is used in a great variety of applications by a large number of tools. Key advantages include the following, (with an X3D overview video linked in Figure 4):

a. X3D is an ISO-ratified file format and run-time architecture to represent and communicate 3D scenes and objects.

b. Multiple open-source X3D players allow integration as part of any Web page without requiring special plugins or prior installations.

c. X3D fully represents diverse forms of 3D data.

d. X3D is developed, maintained, and advanced by the non-profit Web3D Consortium in liaison with multiple other Standard Development Organizations (SDOs).

e. X3D is built on the original Virtual Reality Modeling Language (VRML) standard, still viable since 1997, evolving into the considerably broader ISO X3D standard.

f. X3D provides a system for the storage, retrieval and playback of real-time 3D scenes in multiple applications, all within an open architecture that supports a wide array of domains and user scenarios. X3D model import/export/conversion is widely supported.

Immense detail and specific examples on the use of X3D can be found in Web3D Consortium (n.d.-a, n.d.-b), Havele et al. (2024), and th X3D Models of U.S. Naval Academy Campus (n.d.).

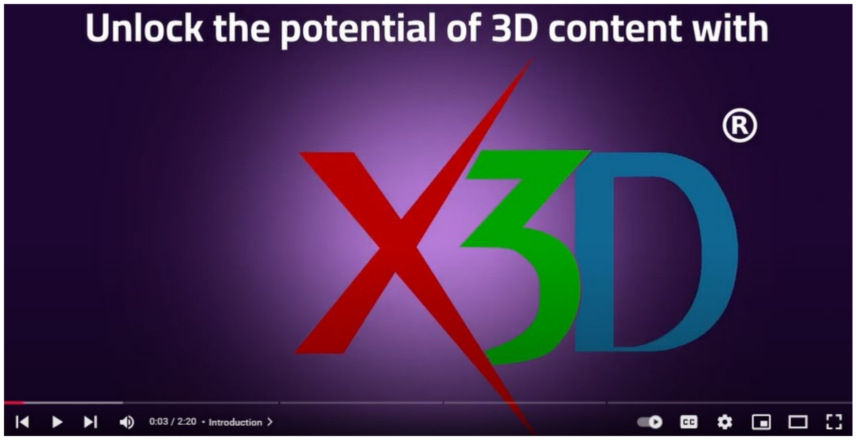


Figure 4. Welcome to X3D Video Shows that X3D Graphics Models are Usable for Multiple Purposes

Of especial interest is that X3D models are portable. X3D Graphics allows modelers to create a Web-based experience, on handhelds or tablets or laptops or big-screen display. This approach is independent of operating systems and reusable by many application libraries. HTML user interfaces can optionally be added to 3D interfaces. Figure 5 illustrates the X3D models for USNA displayed in an immersive environment.

Figure 5. The Virginia Tech (VT) CUBE Walkthrough Video Shows How the Same Unmodified X3D Models Can Also Be Shown in Large-Scale 3D Virtual Environments (3DVEs)

As a potent application for regular collaborative work in 3D environments, the SPIDERS3D collaborative visualization system provides access to shared X3D models to all Navy and USMC personnel, on unclassified networks, using CAC authentication for secure access. A growing X3D model library provides many hundreds of Navy-relevant 3D objects. Multiple locales of interest often include shore-side 3D content of interest, such as pier equipment and exactly spaced pilings retrieved from the official facilities database, calibrated to 1m scale and ready to serve as geospatial context. Figure 6 shows a flexible workflow supporting multiple paths for content preparation and publication, followed by broad subsequent sharing and collaboration. Further details are available in the SPIDERS3D Collaborative Visualization System from NAVFAC/EXWC (n.d.), the SPIDERS3D Program Overview and Collaboration Walkthrough (n.d.), the SPIDERS3D Virtual Sand Table (n.d.), Hall (2024), and Viana et al. (2009).

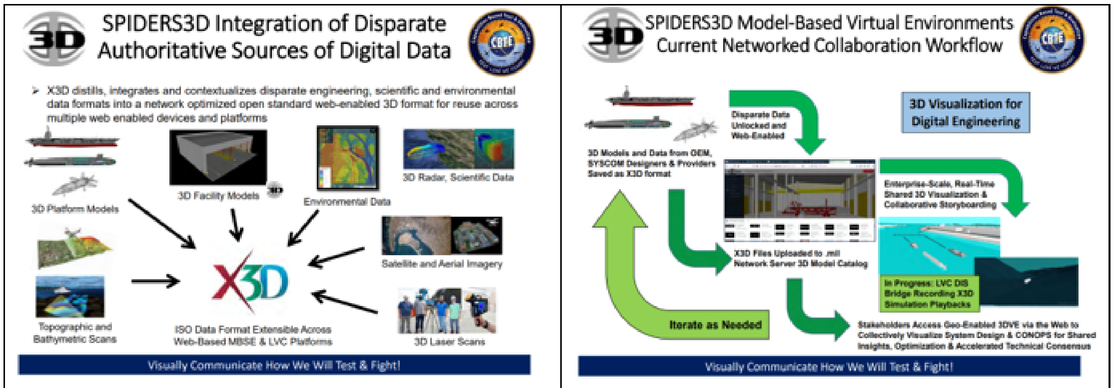


Figure 6. The SPIDERS3D Web-Based Virtual Environment Offers a Collaboration Path Across USN/USMC for 3D Visualization. The SPIDERS3D Overview Presentation and SPIDERS3D Walkthrough Video Describe All System Architectural Elements.

Tools Supporting Model Authoring, Conversions, and Validation

NPS has produced a number of model validation tools that are bundled together in the open-source X3D-Edit authoring tool. This tool suite provides all technical capabilities to perform data ascertainment. Adherence to metadata conventions and canonicalization of text formatting to enable document security and archival publishing. Additional features emphasize model validation, import/export conversions via multiple file-encoding formats, detailed tooltips, reference links, and ancillary software support. Online availability of these open-source assets can be found in Brutzman (n.d.-a, n.d.-b, n.d.c), Brutzman and Daly (2007), *X3D-edit open-source modeling tool*. (n.d.), and Brutzman and Puk (n.d.).

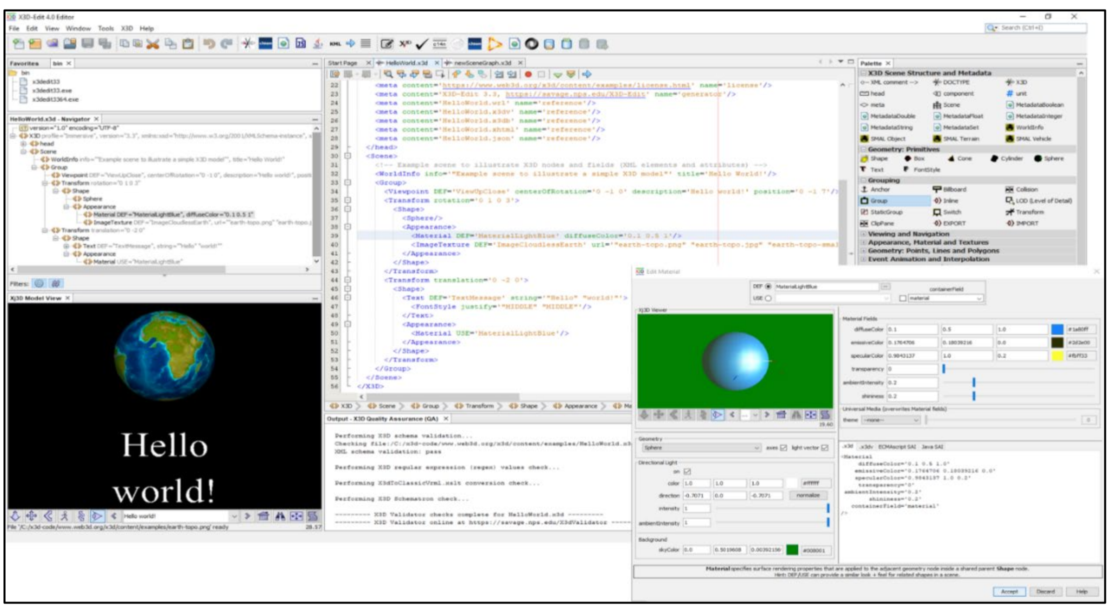


Figure 7. X3D-Edit is a Free, Open-Source X3D Graphics Authoring Tool for Simple High-Quality Authoring, Editing, Import/Export, Validation and Viewing of X3D Scenes

Potential Future Benefits to Navy and Marine Corps

The following examples highlight the versatility and value of a highly precise geospatial, web-based rendering of a naval installation. Multiple benefits can enhance various aspects of installation planning, training, education, and exercises.

Installation Planning

Enhanced Visualization: The 3D rendered scene can be geometrically and geospatially accurate, derived from authoritative models and data sources like CAD, Revit, GIS, and imagery. This allows for better visualization and understanding of the installation layout and infrastructure.

Optimized Model File Sizes: The models are optimized for constrained enterprise network bandwidth, making them accessible to anyone with access to the NAVFAC portal. This ensures that even large-scale models can be shared and viewed efficiently.

Sustainable Digital Engineering: Web-based, open standards enable sustainable cross-SYSCOM digital engineering processes across full platform/program lifecycles. This facilitates long-term planning and development.

Training

Real-time Collaboration: Shareable publication and real-time, web-based collaboration of diverse 3D models enable effective system engineering activities, regardless of the original data source. This is particularly useful for training workshops and exercises.

Comprehensive Training Programs: Training workshops can be conducted at various locations, and web-based handbook training courses can be developed to ensure continuity and accessibility of training materials.

Education

Interactive Learning: Geospatial rendering can be used as an educational tool to provide interactive learning experiences. Students and trainees can explore the installation in a virtual environment, enhancing their understanding of the infrastructure and operations.

Climate Resilience Education: The rendering can be used to educate personnel on climate resilience and adaptation planning. For example, the Climate Action Plan includes comprehensive installation assessments and resilience planning activities that can be tested and evaluated by USNA's Center for Energy Security and Infrastructure Resilience (CESIR).

Exercises

Scenario Planning: The geospatial rendering can be used to simulate various scenarios for exercises, allowing personnel to practice and prepare for different situations. This can include emergency response, security drills, and operational planning.

Performance Tracking: The rendering can be used to track progress and performance during exercises, ensuring that installation readiness is tested and improved continuously.

Strengths Weaknesses Opportunities Threats (SWOT) Analysis

The many useful capabilities shown by the SPIDERS3D system have been hard-earned through years of effort. Some problems (such as misunderstanding Naval needs to share 3D data) remain unresolved. The following strengths, weakneses, opportunities, and threats (SWOT) analysis in Figure 8 summarizes "lessons learned" over the lifetime of this long-running project.

Acquisition Utility of SPIDERS3D, Scenario Planning and Visualization through Archival 3D Model Publication Using the X3D International Standard

Strengths

- Unique Department of Defense (DoD) capability with broad Web access secured via CAC.
- Open standards to archive data assets, sustained by stable improvements over 30 years.
- Growing set of tools, specifications, tutorials, example models with HTML page integration.
- Full interoperability with Web standards, handheld devices, both small and large flat-screen displays, XR interfaces and virtual reality (VR) displays, etc.
- Expected future compatibility with emerging Metaverse Standards Forum (MSF) capabilities via 3D Web Interoperability Working Group.
- Global compatibility with all allied partner nations with added multilingual support possible.

Weaknesses

- DoD and individual services have no comprehensive strategy for 3D data or interoperability.
- DoD provides negligible institutional support for Standards Development Organizations (SDOs) such as Web3D Consortium or World Wide Web Consortium (W3C).
- Data rights are not yet well understood when crafting contracts producing 3D assets.
- Open archival standards and interoperability requirements are not yet well defined or applied during the acquisition process.

Opportunities

- Demonstration projects show significant repeatable value that deserve recurring efforts.
- Web3D Consortium evolves best practices with industry, academia, public, and government.
- It is now possible to define suggested contract requirements, metrics, and acceptance criteria.
- Common solutions benefit both government and all industry partners, enabling long-term cost-reduction progress in combination with greater equipment interoperability

Threats

- Shore-side activities in geospatial planning are often disconnected from operational efforts.
- Industry prefers "lockin" to proprietary commercial products, avoiding data interoperability, requiring regular license renewal, with no guarantee that assets remain usable over time. (Sardonic heuristic: "the best time for an opponent to attack is the day after licenses expire.")
- Supporting government needs, assets and requirements is often considered contrary to long-term corporate strategies, not considered profitable in absence of government requirements and incentives. DoD commands need to own critical 3D assets, not rent them.
- Limited availability and understanding of data interoperability leads to uninformed, mistaken and forced decision making, thus perpetuating counterproductive practices.

Figure 8. Acquisition-Oriented SWOT Analysis Regarding Use of Open Standards for Data Preservation and 3D Model Archiving Using SPIDERS3D and X3D Graphics

Recommendation: 3D Data and Model Review

All relevant X3D model data and raw scans, produced in accordance with Brutzman and Punk (n.d.), are currently held by NAVFAC-tasked contractors. NPS has requested that these important government-funded datasets be delivered to the government. Placing these assets in a trusted, shared location at NPS or USNA for further review by appropriate personnel. Such a review will lead to recommendations for the release of information at either Controlled Unclassified Information (CUI) or unrestricted access. Leadership stakeholders with authority for approving such release include senior personnel at USNA, NPS, NSAA, and NAVFAC/EXWC.

Well-established precedent and practice already exist for such work. For example, the NPS Savage Model Archives are configured for careful handling of both CUI and unrestricted X3D models. Access permissions during the ascertainment can be further granted to USNA faculty, NSAA and NAVFAC personnel, and supporting contractors, as needed. NPS is capable of (and willing to) support hosting these models in a controlled fashion, following all stakeholder guidance, in order to collaboratively perform a security assessment of relevant 3D data and X3D models with other project partners. Successfully executing and documenting such practices holds broader value. Careful attention to detail during this work is likely to further improve candidate best practices and contracts (such as Hall [2024], Viana et al. (2009), and Naval Facilities Engineering Command [2020]) for future Navy/USMC procurement of valid, reusable, standards-based 3D datasets and 3D models.

Next Steps

This work is a multi-year case study resulting in major base improvements. Upon careful review and concurrence regarding what data is releasable, many activities are possible. Future ascertainment activities are interesting and are expected to continue building significant value. Important potential efforts to improve defense acquisition include:

a. Given specific lessons learned (including the incomplete release of X3D assets produced during conduct of Naval Facilities Engineering Command [2020], for example), create specific recommendations for Navy contracting guidelines when procuring and producing 3D models and data assets that deserve either private sharing or public re-use.

b. Addition of USNA campus into SPIDERS3D model archives for CAC-mediated access by personnel across Navy/USMC. Lessons learned can motivate collaborative planning of similar protections for other facilities and installations.

c. Availability for planning faculty and student projects at NPS and USNA.

d. Consider follow-on work regarding climate-change planning patterns for additional bases, possibly hosting a jointly sponsored workshop.

    - USNA Center For Energy Security and Infrastructure Resilience (CESIR), https://www.usna.edu/CESIR
    - NPS Climate and Security Network, https://nps.edu/web/climate-and-security

Acknowledgements

This multidisciplinary effort helps installation resiliency and operational readiness for both Navy and USMC. The author gratefully thanks the following individuals for their contributions to this point paper, ongoing efforts in development of X3D open standards, as well as sustained efforts in the design and production of the USNA campus models.

- Alex Viana, SPIDERS3D Program Manager, Naval Facilities Engineering and Expeditionary Warfare Center (NAVFAC EXWC)
- Dr. Nicholas Polys, Web3D Consortium President and Virginia Tech (VT)
- Vince Stamper, Innovation Manager, Trident Refit Facility (TRF), Bangor Washington
- Terry Norbraten, Dr. Amela Sadagic, NPS Modeling Virtual Environments Simulation (MOVES) Institute and Dr. Jake Didoszak, NPS Mechanical Aeronautic Engineering (MAE) Department
- John Ouellette and Casey Gomez, Versar Inc.

- Michael Russalesi, Tom Cowan, and Adam Barkley, Synergy360 Software Design
- Dr. Richard Puk, Intelligraphics and Web3D Consortium Standards Development
- Anita Havele, Web3D Consortium Executive Director
- Zoe Johnson and Sara Phillips, Naval Support Activity Annapolis (NSAA)
- CDR Kenneth Maroon USN, Ph.D., U.S. Naval Academy (USNA) Computer Science Department
- Robert Lansdale, Chief Technical Officer (CTO), Okino
- Dr. Dale Moore, NPS Department of Defense Management (DDM) and President, Southern Maryland Navy Alliance

References

Brutzman, D. (n.d.-a). *X3D resources*. https://www.web3d.org/x3d/content/examples/X3dResources.html

Brutzman, D. (n.d.-b). *X3D scene authoring hints*. https://www.web3d.org/x3d/content/examples/X3dSceneAuthoringHints.html

Brutzman, D. (n.d.-c). *X3D tooltips*. https://www.web3d.org/x3d/tooltips/X3dTooltips.html

Brutzman, D. & Daly, L. (2007, April). *X3D graphics for web authors*. Morgan Kaufmann Publishers, Elsevier. https://x3dgraphics.com

Brutzman, D. & Puk, R. (n.d.). *X3D standards progress.* https://www.web3d.org/x3d/progress.

Del Toro, C. (2020, May). *Climate Action 2030*. Department of the Navy. https://www.navy.mil/Press-Office/Press-Releases/display-pressreleases/Article/3041221/department-of-the-navy-releases-climate-action-2030, https://www.navy.mil/Portals/1/Documents/Department%20of%20the%20Navy%20Climate%20Action%202030%20220531.pdf

Franchetti, L. (2024). *Chief of Naval Operations, navigation plan for America's warfighting Navy 2024*. Department of the Navy. https://www.navy.mil/Leadership/Chief-of-Naval-Operations/CNO-NAVPLAN-2024

Hall, K. (2024, November). *Digital engineering transformation tactics, techniques, and procedures for shared visualization*. FY24 Naval Innovative Science & Engineering (NISE) Tech Transition, Naval Facilities Engineering Expeditionary Warfare Center (NAVFAC EXWC).

Havele, A., Plesch, A., & Polys, N. (2024, September). Conceptualizing interoperable 3D geospatial data visualization with X3D and OGC 3D tiles. *Web3D 2024 Conference*. https://dl.acm.org/doi/10.1145/3665318.3677167

Naval Facilities Engineering Command. (2020, September 10). *Military installation resilience plan for U.S. Naval Academy* (Firm Fixed Price (FFP) Contract #N6247019D4004).

Security through obscurity. (n.d.). In *Wikipedia*. https://en.wikipedia.org/wiki/Security_through_obscurity

SPIDERS3D collaborative visualization system from NAVFAC/EXWC. (n.d.). Overview summary. https://gitlab.nps.edu/Savage/Spiders3dPublic/-/blob/master/README.md

SPIDERS3D program overview and collaboration walkthrough. (n.d.). https://gitlab.nps.edu/Savage/Spiders3dPublic/-/blob/master/demonstrations/NRWG2020/README.md, https://gitlab.nps.edu/Savage/Spiders3dPublic/-/raw/master/videos/demonstrations/NRWG2020/Spiders3dOverviewWalkthrough.mp4, https://gitlab.nps.edu/Savage/Spiders3dPublic/-/raw/master/presentations/IseaFutureDataScience2022/VianaBrutzmanNSWC_ISEA_SPIDERS3D_14JUN2022.pdf

SPIDERS3D virtual sand table. (n.d.). https://gitlab.nps.edu/Savage/Spiders3dPublic/-/blob/master/demonstrations/VirtualSandTable/README.md

United States Naval Academy. (n.d.). *Social media resources: Find the Naval Academy on Facebook, Twitter, YouTube, Flickr, Instagram and Snapchat*. https://www.usna.edu/Social-Media

United States Naval Academy. (2022, November 23). *Naval Academy breaks ground on sea wall project*. https://www.usna.edu/NewsCenter/2022/11/NAVAL_ACADEMY_BREAKS_GROUND_ON_SEA_WALL_PROJECT.php

United States Naval Academy. (2023, April 27). *USNA, NSAA release installation resiliency plan*. https://www.usna.edu/NewsCenter/2023/04/USNA,_NSAA_RELEASE_INSTALLATION_RESILIENCY_PLAN.php

Viana, A., Brutzman, D., & Polys, N. (2009, December 6). *Collaborative 3D visualization for ashore, afloat and expeditionary readiness workshop*. Virginia Tech Executive Briefing Center. https://www.web3d.org/event/collaborative-3d-visualization-ashore-afloat-and-expeditionary-readiness-workshop

Web3D Consortium. (n.d.-a). *Capabilities*. https://www.web3D.org Web3D Consortium. (n.d.-b). *What is X3D*. https://web3d.org/x3d/what-x3d

X3D-edit open-source modeling tool. (n.d.). https://savage.nps.edu/X3D-Edit.

X3D Models of U.S. Naval Academy Campus for Naval Support Activity Annapolis (NSAA) with Virginia Tech (VT) Cube. (n.d.). Walkthrough video. https://gitlab.nps.edu/Savage/Spiders3dPublic/-/blob/master/demonstrations/AnnapolisUSNA/README.md, https://gitlab.nps.edu/Savage/Spiders3dPublic/-/raw/master/demonstrations/AnnapolisUSNA/NSAA_USNA_3D_Viewer_Demonstration_Video_022522_v5.mp4, https://gitlab.nps.edu/Savage/Spiders3dPublic/-/raw/master/demonstrations/AnnapolisUSNA/VirtuWorlds_NSAA_USNA_Cube.mp4

THIS PAGE INTENTIONALLY LEFT BLANK

Automating AI Expert Consensus: Feasibility of Language Model-Assisted Consensus Methods for Systems Engineering

Ryan Bell—is an 9-year experienced engineer in the defense industry. In his current role at Naval Information Warfare Center Atlantic (NIWC LANT), Ryan 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 life cycle. Ryan earned a BS in electrical engineering from Clemson University and a MS in electrical engineering from Clemson University with a focus on electronics, and 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.bell@nps.edu]

Ryan Longshore—is a 20-year veteran of the defense and electric utility industries. At NIWC LANT, he leads teams developing and integrating new technologies into Navy command centers. He is involved in the Navy's digital engineering transformation with a focus on model-based systems and model-based engineering. Ryan holds a BS in electrical engineering (Clemson) and an MS in systems engineering (SMU), and is pursuing a PhD in systems engineering at the Naval Postgraduate School. He is a South Carolina Registered Professional Engineer and an INCOSE CSEP, and holds the OMG SysML Model Builder Fundamental Certification. [ryan.longshore@nps.edu]

Raymond Madachy, PhD—is a Professor in the Systems Engineering Department at the Naval Postgraduate School. His research interests include systems engineering tool environments for digital engineering, modeling and simulation of systems and software engineering processes, generative AI, and system cost modeling. He has developed cost estimation tools for systems and software engineering, and created the Systems Engineering Library (se-lib). His books include Software Process Dynamics, What Every Engineer Should Know about Modeling and Simulation, What Every Engineer Should Know about Python, and he co-authored Software Cost Estimation with COCOMO II and Software Cost Estimation Metrics Manual for Defense Systems. [rjmadach@nps.edu]

Abstract

Expert consensus is a critical component of decision-making in systems engineering, where stakeholder input and complex trade-offs must be carefully weighed. Traditionally, consensus-building techniques such as the Delphi Method, Nominal Group Technique (NGT), and Multi-Voting have been used to aggregate expert human opinions systematically. Constant lingering challenges prove to be deterrents such as time-intensive and extensive coordination efforts required to gather Subject Matter Experts (SMEs). With the advent of Large Language Models (LLMs), there exists the potential to capture the expert knowledge and leverage AI to streamline consensus-building.

This conceptual paper explores the feasibility of LLM-assisted consensus methods in the context of systems engineering. We evaluate consensus methods based on their structure, expert interaction requirements, and compatibility with LLMs, followed by identifying which methods could be enhanced through AI-driven automation. Through a comparative analysis, we hypothesize the methods best suited for LLM augmentation or full automation and explore their potential applications in systems engineering. Finally, we discuss future research directions for both AI-driven and hybrid human-AI consensus frameworks.

Keywords: Large Language Models (LLMs), Artificial Intelligence (AI), Consensus Methods, Systems Engineering, Feasibility Study

Introduction

The presence of accessible, capable AI systems has become widespread and presents a tool that should be leveraged intelligently as a force multiplier. The next generation of language models will require a shift from a "one size fits all" model to domain-specific models (Ling et al., 2024). The models can be trained on their own or fine-tuned from foundational models. Foundational models are models trained on general knowledge. Conceptually, the research discussed herein focus on the interactions between and how to employ these domain-specific models. With each model trained on domain-specific knowledge, the similarity to being considered “AI Subject Matter Experts (SMEs)” as the same as having “human SMEs” starts to come to fruition.

The interaction between SMEs is a commonly orchestrated event for systems engineers. Systems engineers, acting as the glue between SMEs, sponsors, and project managers, are well poised to leverage domain-specific models in situations where a SME may not be available or too costly. This paper will explore the consensus methods commonly used by systems engineers for soliciting domain-specific knowledge to make informed decisions, discuss implementation architectures that are feasible for usage with language models, and propose systems engineering use cases and examine their challenges to implementation.

Overview of Consensus Methods

While language models have an inherent ability to synthesize large corpuses of information, their ability to come to a consensus among several models has been less studied, although interesting effects have been found at scale (Marzo et al., 2025). Some research has been done on hybrid consensus methods, including both humans and AI to come to a consensus (Chen et al., 2023; Fogliato et al., 2022; Hirosawa et al., 2024; Papakonstantinou et al., 2025; Punzi et al., 2024). Research has also been done on some of the challenges associated with hybrid consensus methods (Vaccaro et al., 2024).

The consensus methods considered are among some of the most common, including: the Delphi Method, the Fuzzy Delphi Method, Structured Expert Judgment (SEJ) also called Cooke’s Method, Nominal Group Technique (NGT), the Stepladder Technique, Dialectical Inquiry, and Multi-Voting (Dot Voting). A summary table of each of these consensus methods’ strengths and weaknesses is in Table 1, Consensus Methods Strengths and Weaknesses. A deeper dive into each consensus method follows. The sequence diagrams generated are intended to be representative of the most implementations of each technique, although there were slight variations present between different pieces of literature.

Table 1. Consensus Methods Strengths and Weaknesses

Method	Strengths	Weaknesses
Delphi Method	Reduces bias, allows for geographic dispersion, and provides a systematic approach to achieving consensus.	Time-consuming, lacks interaction, and may not achieve consensus.
Fuzzy Delphi Method	Captures ambiguity and uncertainty in expert opinions.	Complex for non-experts and requires fuzzy logic expertise.
Structured Expert Judgment	Provides quantitative outputs and handles uncertainty effectively.	Resource-intensive and requires expertise.
Nominal Group Technique	Encourages equal participation and produces clear prioritization.	Limited to small groups and time-consuming.
Multi-Voting	Quick, easy to implement, and provides clear prioritization.	May not capture nuances and can be influenced by voting strategies.
Stepladder Technique	Reduces dominance and improves decision quality.	Time-consuming and requires planning.
Dialectical Inquiry	Encourages critical thinking and creative solutions.	Contentious and may not achieve consensus.

The Delphi method is a structured, iterative process used to gather and consolidate expert opinions. It involves multiple rounds of questionnaires, with feedback provided to participants between rounds to encourage convergence of opinions. Experts respond to questionnaires in multiple rounds, with anonymous feedback usually in the form of the group average provided after each round. Consensus is typically defined as a percentage of agreement (e.g., 70–80%) or convergence variance of responses (e.g., +/- 1 on a ranking scale). While this method attempts to reduce bias from dominant personalities through anonymous responses and the use of a facilitator, the process can be time-consuming and may not always achieve the set consensus threshold.

The authors have employed the Delphi method for estimating systems engineering cost model parameters for using AI (Madachy et al., 2025). Previously the Constructive Systems Engineering Cost Model (COSYSMO) and Constructive Cost Model (COCOMO) for software development were developed and calibrated with both expert judgment data via Delphi surveys and historical project data (Boehm et al., 2000; Valerdi, 2005). The Delphi method is also commonly used in clinical settings, among other domain-specific fields (Chan, 2022; Erffmeyer, 1981; Hutchings et al., 2006; Kauppi et al., 2023; Papakonstantinou et al., 2025; Spranger & Niederberger, 2025; Vedantham et al., 2023). The process is captured in Figure 1, Delphi Method Sequence Diagram.

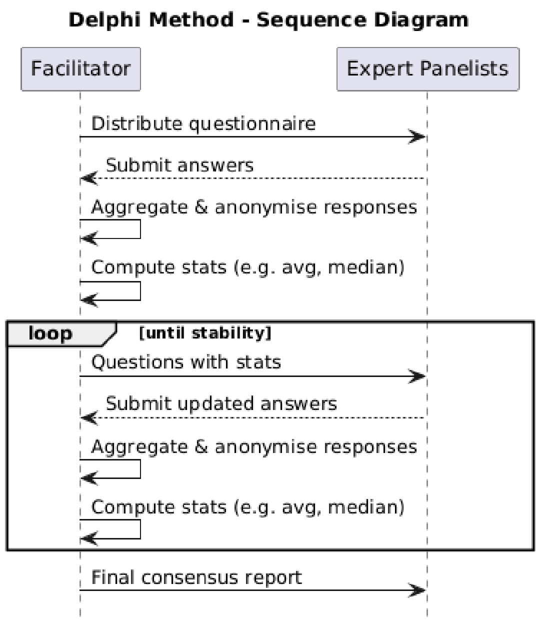


Figure 1. Delphi Method Sequence Diagram

The Fuzzy Delphi method integrates fuzzy logic with the traditional Delphi Method to capture the uncertainty in expert judgments. It is a structured, iterative process used to gather and consolidate expert opinions. It involves multiple rounds of questionnaires, with feedback provided to participants between rounds to encourage convergence of opinions. Experts still respond to questionnaires in multiple rounds, with anonymous feedback provided after each round, but use ranges—a fuzzy score—to compute convergence. This method lends itself best to situations where precise data is unavailable, but the learning curve is steep for facilitators new to fuzzy logic. The Fuzzy Delphi method is commonly used in situations where there is substantial ambiguity that needs to be quantified (Mohamad et al., 2015; Nayebpour & Sehhat, 2023; Padzil et al., 2021; Rahman & Kamauzaman, 2022; Rani et al., 2023). The process is captured in Figure 2, Fuzzy Delphi Method Sequence Diagram.

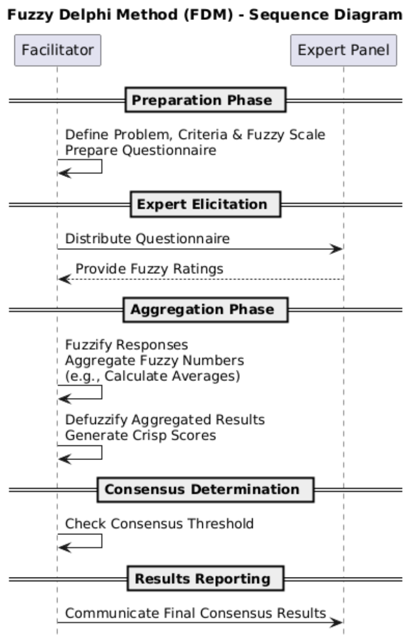


Figure 2. Fuzzy Delphi Method Sequence Diagram

The Structured Expert Judge (SEJ) method, also known as Cooke's method, uses expert opinions to quantify and produce probabilistic estimates. Each expert provides their individual assessment, all responses are aggregated, statistical weighting models are applied, and calibration is included if necessary. The quantitative output of the process is desirable, particularly for ambiguous and complex issues, although the process requires an expert to design. Cooke's method is commonly used within the nuclear field, ecosystems, and public health, among others (Colson & Cooke, 2018; Cooke et al., 2021; Felfernig & Le, 2023; Ullrika Sahlin, 2023). The process is captured in Figure 3, Structured Expert Judgment Sequence Diagram.

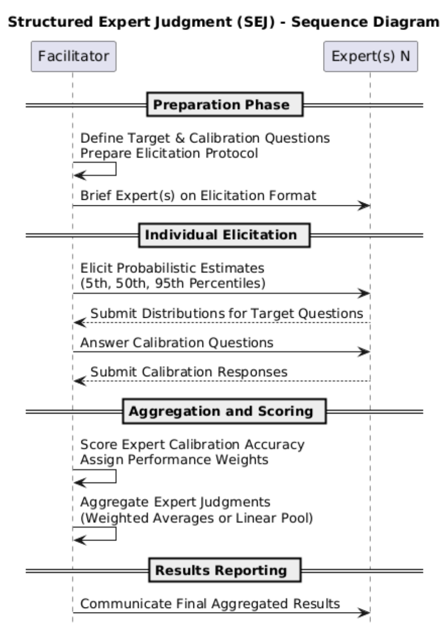


Figure 3. Structured Expert Judgment Sequence Diagram

The Nominal Group Technique (NGT) method is structured as a face-to-face consensus method that combines individual brainstorming with group discussion. NGT is primarily for structured idea generation and prioritization. Individuals brainstorm ideas, then ideas are then shared by each individual, one at a time while ideas are publicly recorded. Once all individuals have shared their ideas, the floor is open to group discussion with the focus on clarification of the ideas. Once clarifications are complete, everyone ranks the ideas that are most important or relevant. With humans, the process is typically limited to small groups, requires a facilitator, and can be time consuming. The output of this process is a list of ranked ideas, which can be used as inputs to other consensus methods to narrow down the list, such as the multi-voting method. Common usages include clinical studies and teaching, among others (Erffmeyer, 1981; Hutchings et al., 2006; Mousa et al., 2022; Rahman & Kamauzaman, 2022). The process is captured in Figure 4, Nominal Group Technique Sequence Diagram.

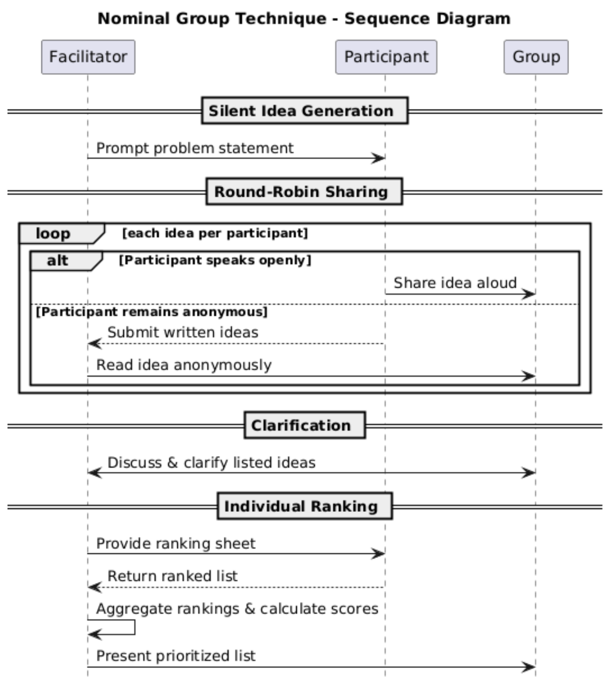


Figure 4. Nominal Group Technique Sequence Diagram

The Stepladder Technique is structured such that individual opinions are gradually added to the group discussion. In a tiered fashion, group sizes gradually increase. The process would start with individuals paired up who discuss their thoughts, followed by merging pairs to form small groups. Discussions continue. Small groups are then merged into a larger group. The process continues until all participants are in a single group. This method typically encourages participation from all members and reduces group think but requires a significant amount of time and is not typically suited for large groups (Rogelberg et al., 1992; Rogelberg & O'Connor, 1998).

The sequence diagram in Figure 5, Stepladder Sequence Diagram, presents a maximum of eight participants, although it could have as many as the facilitator or consensus designer would like and is merely a medium to communicate the process.

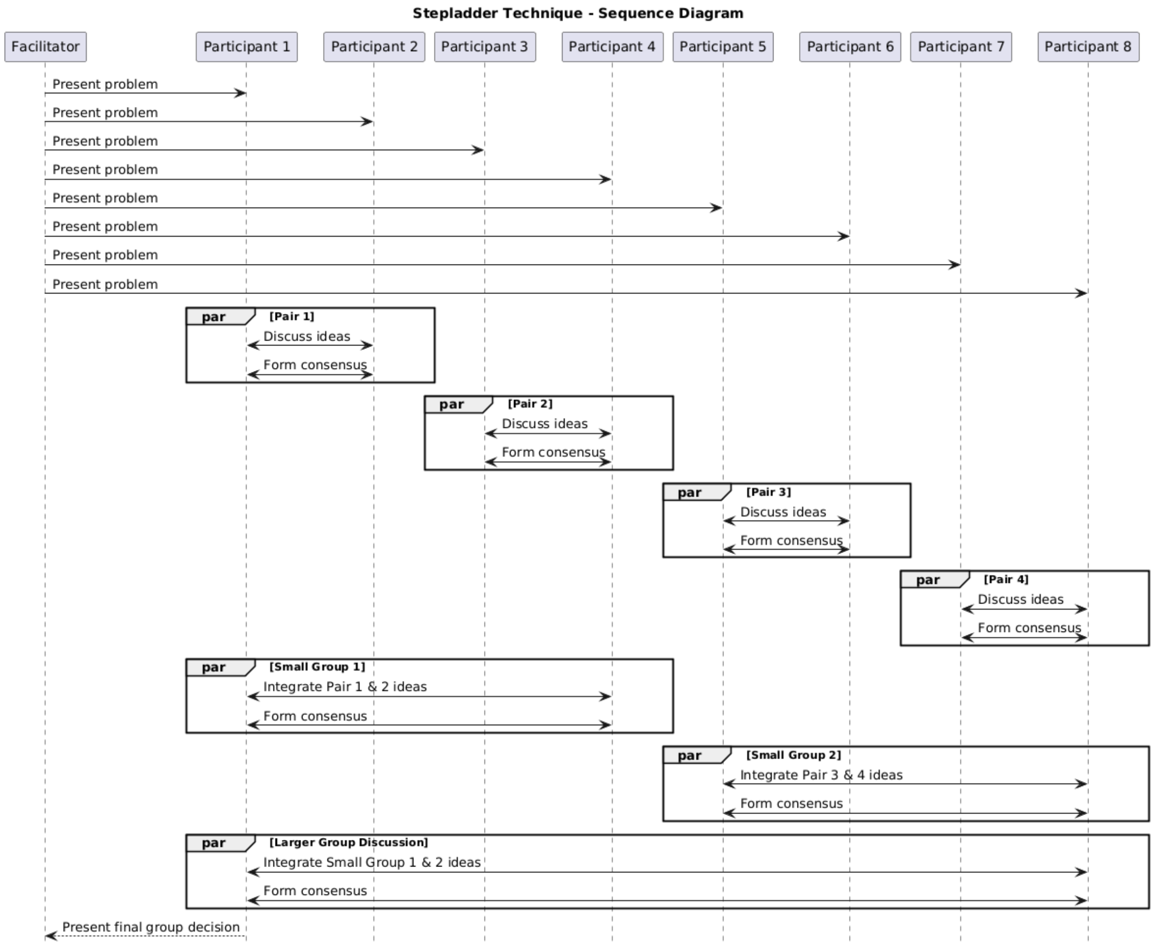


Figure 5. Stepladder Sequence Diagram

The Dialectical Inquiry method is focused on the premise of presenting opposing viewpoints to stimulate critical thinking among a group. Participants are divided up into groups that are assigned to argue for or against a proposition, and the debate continues until a consensus is reached. While the method can encourage creative solutions and foster new viewpoints, it can be contentious, requiring tactful facilitation, or, just as likely to ultimately not reach a consensus (Fjermestad, 1994; Priem & Price, 1991; Tung & Quaddus, 2001). The process is captured in Figure 6, Dialectical Inquiry Sequence Diagram.

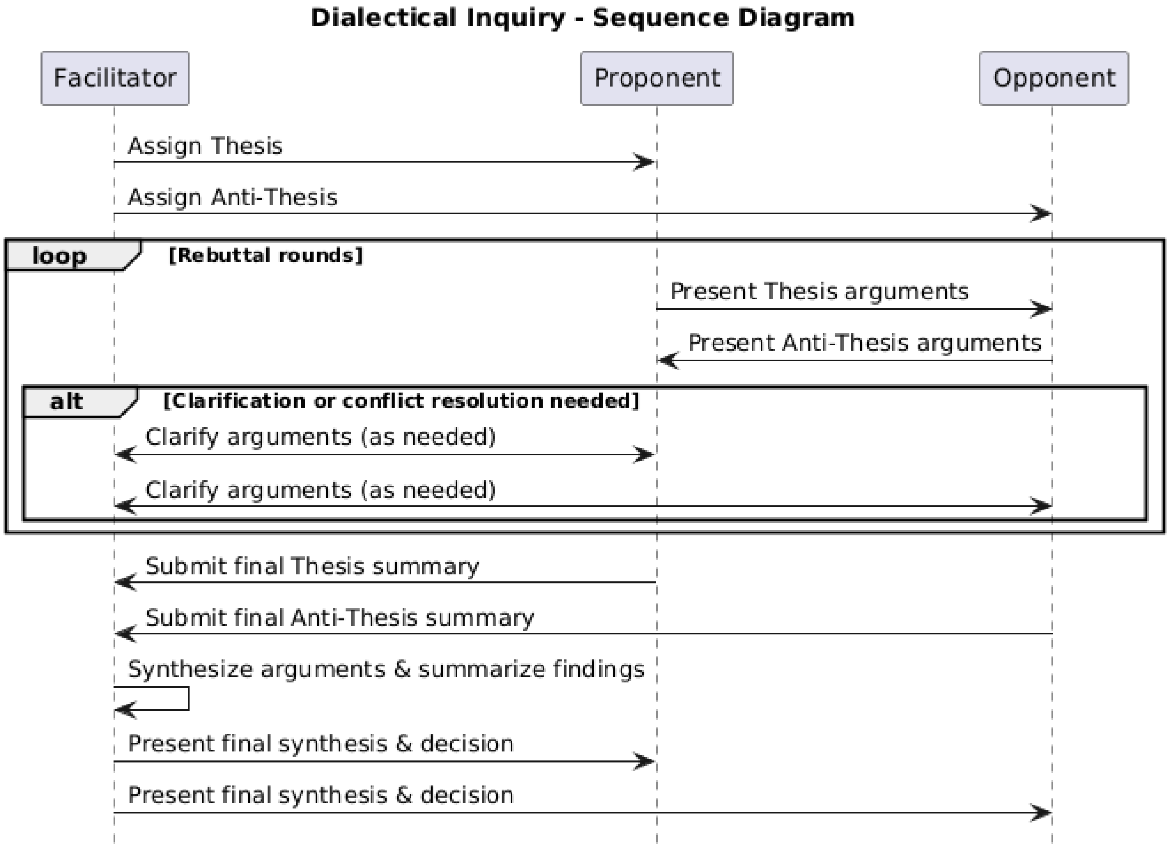


Figure 6. Dialectical Inquiry Sequence Diagram

The Multi-Voting method, also known as Dot Voting, is commonly used in the six-sigma process, where individuals vote on multiple items from a list. The list is generated ahead of time and may be generated via another consensus method like the NGT method. Multi-voting is then used to narrow down the list of options based on the group's consensus. Participants allocate their votes by either assigning a limited number (usually half the number of items) or by ranking all items on the list. The votes are then compiled and top items are presented. Multi-voting weighs every individual's vote equally. The selection process can become time-consuming or cumbersome for large lists. Multi-voting is used to narrow down a list of options as it is a simple voting mechanism used in Six Sigma practices and a variety of fields (American Society for Quality, 2025; Atlassian Community, 2024; Digital Healthcare Research, 2025; Hessing, 2015; Nielsen Norman Group, 2025). The process is captured in Figure 7, Multi-Voting Sequence Diagram.

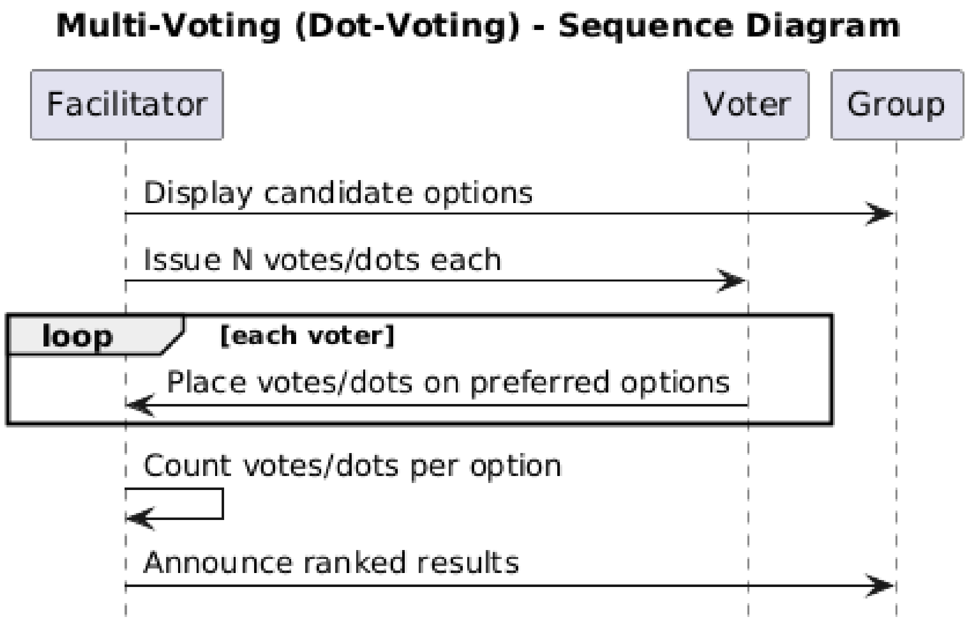


Figure 7. Multi-Voting Sequence Diagram

Summary of Consensus Method Characteristics

For each consensus method, a few characteristics were captured to support decision making for consensus method selection. The characteristics chosen are the columns found in Table 2, Consensus Methods Overview, and are anonymity, iteration, facilitation, output type, group interaction type, and the aggregation method.

Anonymity indicates whether participants provide input anonymously, which can affect group dynamics and bias mitigation. Possible values for this field include Yes, No, Partial, or Optional.

- Yes means full anonymity is maintained between participants.
- No means contributions are made openly.
- Partial means some anonymity exists during one or more of the stages of the consensus process.
- Optional means anonymity may or may not be used depending on the implementation style.

Iteration indicates whether the method includes repeated rounds of input and feedback, which can help refine judgments and converge on consensus. Possible values for this field include Single Round, Multiple Rounds, Built-in, or Optional.

- Single Round means the method is conducted in a single structured session without repetition.
- Multiple Rounds means the method explicitly involves repeated cycles of input, feedback, and revision.
- Built-in means iterative progression is inherently embedded in the method's structure.
- Optional means iteration is not required but can be included at the facilitator's discretion or based on group needs.

Facilitation refers to the level of structured guidance needed to execute the method effectively. Possible values for this field include Facilitator-Driven or Facilitator-Supported.

- Facilitator-Driven means a central facilitator is required to guide the process, manage feedback rounds, and enforce structure.
- Facilitator-Supported means a facilitator helps organize and maintain flow but does not drive every part of the process.

Output Type describes the nature of the results produced by the method. This determines whether the outcomes are narrative, numerical, or both, which influences how results are interpreted and used in decision-making. Possible values for this field include Qualitative, Quantitative, or Both.

- Qualitative means the outputs are primarily textual in nature.
- Quantitative means the outputs are numerical, such as rankings, vote tallies, or probabilistic scores.
- Both means the method can produce either qualitative insights or quantitative metrics.

Group Interaction Type identifies how participants communicate and collaborate during the method, which affects scheduling, group dynamics, and tool selection. Possible values for this field include Asynchronous or Synchronous.

- Asynchronous means participants provide input independently and at different times.
- Synchronous means participants interact in real time.

Decision Aggregation Method defines how individual participant inputs are synthesized into a collective judgment. This mechanism is central to reaching consensus or selecting preferred alternatives. Possible values for this field include None, Optional, Ranking, Count-Based, Scoring, or Fuzzy Scoring.

- None means there is no formal aggregation; consensus may emerge through discussion or argumentation.
- Optional means aggregation may or may not be used depending on context or facilitation style. The aggregation method could be one of the other methods but is not required.
- Ranking means participants order alternatives by preference, typically in descending importance.
- Count-based means options are selected or voted on with multiple tallies.
- Scoring means participants assign numeric ratings to options, which are then averaged or aggregated.
- Fuzzy Scoring means participants express uncertainty through ranges or fuzzy values (e.g., minimum, most likely, and maximum), which are aggregated using fuzzy logic methods.

Table 2. Consensus Methods Overview

Method	Anonymity	Iteration	Facilitation	Output Type	Group Interaction Type	Aggregation Method
Delphi	Yes	Multiple Rounds	Facilitator-Driven	Both	Asynchronous	Scoring
Fuzzy Delphi	Yes	Multiple Rounds	Facilitator-Driven	Quantitative	Asynchronous	Fuzzy Scoring
Structured Expert Judgment	Yes	Optional	Facilitator-Driven	Quantitative	Asynchronous	Scoring
Nominal Group Technique	Partial	Single Round	Facilitator-Driven	Both	Synchronous	Ranking
Multi-Voting	Optional	Optional	Facilitator-Supported	Quantitative	Synchronous	Count-based
Stepladder Technique	No	Built-in	Facilitator-Supported	Both	Both	Optional
Dialectical Inquiry	No	Multiple Rounds	Facilitator-Supported	Qualitative	Synchronous	None

The Implementation of Consensus Methods with Language Models

In order to understand the trade-offs between implementing consensus methods with language models, this research proposes criteria to qualitatively assess between the methods. A list of criteria was brainstormed to include: parallelizability, number of personas, agent persona archetypes, inter-agent communication pattern, and memory length.

The parallelizability criterion is how much of the method can be parallelized (e.g., agents working independently at the same time) where: High is fully parallel, Medium is some steps parallel, some sequential, and Low is mostly sequential. The number of AI personas is the recommended minimum number of distinct AI agents needed to implement the method. Agent persona archetypes are the types of roles or behavioral archetypes needed among the AI agents. The inter-agent communication pattern is how the AI agents exchange information during the process. Last but not least, memory length refers to how much dialogue or context history each agent needs to maintain during the method's execution where a single chat only requires one-off responses and conversational requires ongoing memory of turns or rounds.

Table 3, Consensus Method Implementation Characteristics, summarizes all of the evaluated criteria for each consensus method. In circumstances where a synthesizer persona is recommended, the role can be typically merged with the facilitator role, which is synonymous with the sequence diagrams. Some assumptions were made, including: 1. This is the logical formation of personas, but may be implemented as separate LLM calls or a single LLM stepping through roles and 2. If multi-round option is selected, this would be conversational.

Table 7 Consensus Method Implementation Characteristics

Consensus Method	Parallelizability	# AI Personas	Agent Persona Archetypes	Inter-Agent Communication Pattern[1]	Memory Length
Delphi Method	Medium	3+	Facilitator, Expert, Synthesizer	Hub-and-Spoke	Conversational
Fuzzy Delphi Method	Medium	3+	Facilitator, Expert, Synthesizer	Hub-and-Spoke	Conversational
Structured Expert Judgment	High	3+	Facilitator, Expert, Synthesizer	Hub-and-Spoke	Single Chat[2]
Nominal Group Technique	Low	4+	Facilitator, Creative Expert, Reasoning Expert, Summarizer	Group Broadcast	Conversational
Multi-Voting	High	2+	Faciliator, Participant	Blind Broadcast	Single Chat
Stepladder Technique	Medium	3+	Faciliator, Participant, Synthesizer	Progressive Entry	Conversational
Dialectical Inquiry	Low	3+	Facilitator, Thesis Supporter, Antitheses Supporter	Sequential Debate	Conversational

The agent persona archetypes are major roles, including the facilitator, expert, synthesizer, creative expert, reasoning expert, participant, thesis supporter, and antithesis supporter. Some of these personas could be merged under certain circumstances, like the facilitator and synthesizer. In general, the following purpose of each of these archetypes is

- Facilitator: Guides the process, enforces rules, moderates the flow. Also known as the Conductor within agentic frameworks.

- Expert: Provides substantive technical input or judgment. A generalization of a creative or reasoning expert.
- Synthesizer: A decomposition of the facilitator role to summarize and aggregate information.
- Creative Expert: An expert that focuses on brainstorming new ideas in early stages.
- Reasoning Expert: An expert that focuses on substantiating, prioritizing, or ranking options.
- Participant: A general contributor without major specialization.
- Thesis Supporter: Defends an assigned position.
- Antithesis Supporter: Critiques the thesis with counter-arguments.

Each of these persona archetypes generally values a different level on the "creativity" scale, which is synonymous with temperature for language models. The relationship between temperature and persona archetype is continued in Table 4, Agent Persona Temperatures.

Table 4. Agent Persona Temperatures

Persona Archetype	Suggested Temperature	Rationale
Facilitator	Low	Must keep structure, restate prompts, and remain neutral. Deterministic output avoids accidental bias or drift.
Expert	Low–Medium	Needs factual depth with little room for nuance or hypothesis generation. Too much randomness risks misinformation; too little may freeze creative problem solving.
Synthesizer / Summarizer	Low	Primary duty is faithful condensation. Higher temperature could invent facts or reorder logic.
Creative Expert	Medium–High	Charged with idea generation. Higher temperature encourages novel alternatives and divergent thinking.
Reasoning Expert	Low–Medium	Focus on logical evaluation; moderate temperature keeps reasoning flexible but still disciplined.
Participant	Low–Medium	Casting or explaining a preference benefits from mild variability (tie break rationales) but must stay consistent with criteria.
Thesis Supporter / Antithesis Supporter	Medium–High	Goal is vigorous argumentation. Higher temperature produces persuasive rhetoric, counter examples, and creative rebuttals, effectively fueling dialectical tension.

The characteristics identified and qualities assessed will help with adaptation into agentic frameworks like CrewAI, Autogen, OpenAI's swarm, among many others from the open source community (GitHub, 2024; Microsoft, 2023; n8n.io, 2025; OpenAI, 2024; SuperAGI, 2025).

Systems Engineering Applications of LLM-Centric Consensus Methods

Consensus plays a critical role in systems engineering by ensuring that the boundaries of complex technical trade spaces reflect the collective judgment of multidisciplinary stakeholders such that the systems engineer can make an informed decision. In some cases, expert consensus enables the reconciliation of conflicting priorities—such as cost, performance, schedule, and safety—through structured deliberation while in other cases, consensus simply populates the bounds of the trade space. As systems grow in complexity with exponential interdependencies, the ability to achieve consensus among domain-specific SMEs becomes a cornerstone of successful systems engineering practices.

Language models are emerging as powerful AI tools for decision support in systems engineering. Given their capability for synthesizing large swaths of information and offering structured insights, they are a natural support tool for systems engineers. When integrated into tools like Model-Based Systems Engineering (MBSE) environments with SysMLv2 textual notation, LLMs can be just as aware as the systems engineer, with hopes of enhancing traceability by cross-referencing system artifacts. In the Department of Defense (DoD), domain-specific knowledge bases are plentiful. Connecting a language model to these domain-specific knowledge bases and simply having the models interact rather than the full exchange of data is desirable for compartmentalization reasoning and security. The ability to have these domain-specific aware, black box models and having an interplay between them may bring a level of consensus on complex topics not before made. To further elaborate on the application of consensus methods in systems engineering, two use cases were chosen to pontificate on how these LLM-centric methods would apply to common systems engineering problems.

The first use case presented is one pertaining to requirements engineering, specifically stakeholder requirements solicitation—arguably the most important stage—where we can use several AI agents with varying personas to brainstorm pertinent stakeholder needs to requirements, followed by a consensus method for pruning this large list into a pruned prioritized list of stakeholder requirements.

We hypothesize that using NGT or the stepladder technique for an initial pass at requirements are both good approaches. In this specific case, we propose that NGT is used for its ability to brainstorm from many viewpoints, followed by an optional multi-voting method for pruning the list should the list be too long. The final pruning would need to be guided by a human, but the facilitation can still occur from an AI persona. This scenario is about surfacing what matters to diverse users. The goal is to ensure each voice is heard and that the initial capability list is representative, even if imperfect or perhaps lacking technical rigor, depending on the personas selected. Personas might include all typical SMEs from an Integrated Product Team (IPT), such as but not limited to mechanical, electrical, structural, aerospace, logistics, and program management.

The second use case presented is about performing risk analysis. Generally, risk analysis requires somewhat specialized knowledge like fault tree analyses, FMEA, or FMECA. In this scenario, it is assumed that the brainstorming phase has been conducted and the focus is on narrowing down the most plausible solution(s).

We hypothesize that given the typically required specialized knowledge, consensus methods that leverage experts like SEJ or a Delphi approach are appropriate. Both methods enable the voice of the experts to be heard with optional iterative feedback between experts. Using the knowledge available to the agents, the question posed would be for a ranking of severity. Personas might include a level of sub-field specificity like reliability engineers or availability engineers instead of more general personas like mechanical, electrical, or similar.

Challenges to Implementation

There are two main categories of challenges to implementation: challenges that are inherent to LLMs in general and challenges that are inherent to the consensus framework used. Challenges like accuracy, bias, role consistency, human diversity at the single model level, and memory are all LLM-inherent challenges with AI technology. When it comes to implementing the consensus framework, things like emulating human diversity where at the group level emergent behavior could be present, iteration management, human oversight of the process, and consensus metrics are all present. The challenges are summarized in Table 5, Challenges to Implementation.

Table 5. Challenges to Implementation

Challenge Category	Challenge	Considerations
LLM-Inherent	Accuracy and bias	RAG, Fine-tuning, benchmarked models
LLM-Inherent	Role consistency	Role templates, temperature tuning
Both	Emulating human diversity	Model/temperature mixing, diverse personas, prompt engineering
Both	Memory and iteration management	Long-context models, vector stores, iterative summarization
Consensus Framework	Human oversight	Human-on-the-loop, checkpoints, audit logs
Consensus Framework	Consensus and convergence metrics	Entropy/rubric metrics, semantic convergence checks, human final judgment

Future Work and Research Opportunities

There is a plethora of future research available in this area—the most obvious is the implementation of each of the consensus methods discussed herein into open source tooling, such as a Python library to be used with language models. The library could include human in the loop versus human on the loop interfaces as well as support for hybrid consensus framework structures for human-AI teaming.

The application of any of these consensus frameworks with AI to a typical systems engineering use case as discussed would inform practitioners about the usefulness of using AI as a force multiplier. Multi-modal or vision-models could be used to assess prototype photographs of systems. Convergence rates for varying temperature or “creativity” levels could be investigated. The ability to scale participants to levels incapable of human participation also warrants investigation, with the hopes of finding emergent behavior not previously possible at scale. For example, the application of large scale ranking without reaching cognitive overload of participants with 100s of items to prioritize.

Conclusion

The usage of consensus methods with AI necessitates further research. The systems engineering field would benefit greatly from gathering consensus from multiple language models across different phases of the systems engineering life cycle. Practicing systems engineers and SMEs could supplement their own knowledge bases with AI personas to enhance viewing problems from different perspectives.

As AI tool suites continue to propagate, systems engineers need to consider the assimilation of AI tools with classical methods of reaching decisions via consensus methods with experts. The ability to process, understand, and make informed decisions within a trade space is only going to become more challenging for systems engineers as systems of systems continue to become more complex. Domain-specific AI models can help relieve some of the complexity—from understanding an entire model-based systems engineering (MBSE) model in SysMLv2 textual form to understanding entire knowledge bases of domain-specific data, gathered from decades of expert practice in the field.

The future of systems engineering will not depend solely on more powerful AI models, but on how effectively humans and machines collaborate. The challenge lies in engineering decision frameworks that balance trust, skepticism, and synthesis across diverse AI and human perspectives for effective, practical implementation.

Acknowledgements

This work has benefited from the use of generative AI tools, including ChatGPT, Claude, and SciSpace, for brainstorming, writing assistance, and code development (Anthropic AI, 2025; OpenAI, 2025; SciSpace, 2025). These tools were employed to enhance conceptual clarity, improve code efficiency, and support literature synthesis. All AI-generated outputs were critically reviewed, refined, and validated to ensure accuracy and alignment with academic integrity. Their contributions were limited to supporting the research process, and final responsibility for the content, analysis, and conclusions remains with the authors.

References

American Society for Quality. (2025). *What is multivoting? NGT voting, nominal prioritization*. https://asq.org/quality-resources/multivoting?srsltid=AfmBOoqNNFsEpllk7pjC74Reu9AhCE_GBNM1Fb8U1wPFElgWNi9yXlKC

Anthropic AI. (2025). *Anthropic home*. https://www.anthropic.com/

Atlassian Community. (2024, March 27). *Dot voting: An effective way for group decision-making*. https://community.atlassian.com/forums/App-Central-articles/Dot-Voting-An-Effective-Way-for-Group-Decision-Making/ba-p/2653536

Boehm, B., Abts, C., & Chulani, S. (2000). Software development cost estimation approaches – A survey. *Annals of Software Engineering*, *10*(1–4), 177–205. https://doi.org/10.1023/A:1018991717352

Chan, P. (2022). An empirical study on data validation methods of delphi and general consensus. *Data*, *7*(2), 18. https://doi.org/10.3390/data7020018

Chen, X., Xu, X., Pan, B., & Zhang, W. (2023). *Human-machine collaboration-driven consensus method for large-scale group decision-making based on reinforcement learning algorithm and its application*. https://doi.org/10.2139/ssrn.4623328

Colson, A. R., & Cooke, R. M. (2018). Expert elicitation: Using the classical model to validate experts' judgments. *Review of Environmental Economics and Policy*, *12*(1), 113–132. https://doi.org/10.1093/reep/rex022

Cooke, R. M., Marti, D., & Mazzuchi, T. (2021). Expert forecasting with and without uncertainty quantification and weighting: What do the data say? *International Journal of Forecasting*, *37*(1), 378–387. https://doi.org/10.1016/j.ijforecast.2020.06.007

Digital Healthcare Research. (2025). *Multivoting*. https://digital.ahrq.gov/health-it-tools-and-resources/evaluation-resources/workflow-assessment-health-it-toolkit/all-workflow-tools/multivoting

Erffmeyer, R. (1981). *Decision-making formats: A comparison on an evaluative task of interacting groups, consensus groups, the nominal group technique, and the Delphi technique.* [Doctor of Philosophy, Louisiana State University and Agricultural & Mechanical College]. https://doi.org/10.31390/gradschool_disstheses.3593

Felfernig, A., & Le, V. M. (2023). An overview of consensus models for group decision-making and group recommender systems. *User Modeling and User-Adapted Interaction*. https://doi.org/10.1007/s11257-023-09380-z

Fjermestad, J. L. (1994). *Group strategic decision-making in a computer-mediated-communications environment: A comparison of dialectical inquiry and constructive consensus approaches*.

Fogliato, R., Chappidi, S., Lungren, M. P., Fitzke, M., Parkinson, M., Wilson, D. U., Fisher, P., Horvitz, E., Inkpen, K., & Nushi, B. (2022). Who goes first? Influences of human-AI workflow on decision making in clinical imaging. *2022 ACM Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3531146.3533193

GitHub. (2024). *CrewAI*. https://github.com/crewAIInc/crewAI

Hessing, T. (2015, November 21). Multivoting. *Six Sigma Study Guide*. https://sixsigmastudyguide.com/multivoting/

Hirosawa, T., Shiraishi, T., Hayashi, A., Fujii, Y., Harada, T., & Shimizu, T. (2024). Adapting artificial intelligence concepts to enhance clinical decision-making: A hybrid intelligence framework. *International Journal of General Medicine*, *17*, 5417–5422. https://doi.org/10.2147/ijgm.s497753

Hutchings, A., Raine, R., Sanderson, C., & Black, N. (2006). A comparison of formal consensus methods used for developing clinical guidelines. *Journal of Health Services Research & Policy*, *11*(4), 218–224. https://doi.org/10.1258/135581906778476553

Kauppi, K., Roos, E., Borg, P., & Torkki, P. (2023). Building consensus on domains of wellness using Finnish and international expert panels: A Delphi-method study. *American Journal of Health Promotion*, 8901171231204147. https://doi.org/10.1177/08901171231204147

Ling, C., Zhao, X., Lu, J., Deng, C., Zheng, C., Wang, J., Chowdhury, T., Li, Y., Cui, H., Zhang, X., Zhao, T., Panalkar, A., Mehta, D., Pasquali, S., Cheng, W., Wang, H., Liu, Y., Chen, Z., Chen, H., … Zhao, L. (2024). *Domain specialization as the key to make large language models disruptive: A comprehensive survey* (No. arXiv:2305.18703). arXiv. https://doi.org/10.48550/arXiv.2305.18703

Madachy, R., Bell, R., & Longshore, R. (2025). A generative AI-driven systems engineering maturity and cost modeling framework. *Proceedings of the 2025 Conference on Systems Engineering Research*.

Marzo, G. D., Castellano, C., & Garcia, D. (2025). *AI agents can coordinate beyond human scale* (No. arXiv:2409.02822). arXiv. https://doi.org/10.48550/arXiv.2409.02822

Microsoft. (2023). *AutoGen*. GitHub. https://github.com/microsoft/autogen

Mohamad, S. N. A., Embi, M., & Nordin, N. (2015). Determining e-Portfolio elements in learning process using fuzzy Delphi analysis. *International Education Studies*, *8*, 171–171. https://doi.org/10.5539/ies.v8n9p171

Mousa, M., Teede, H. J., Garth, B., Winship, I., Prado, L., & Boyle, J. (2022). Using a modified Delphi approach and nominal group technique for organisational priority setting of evidence-based interventions that advance women in healthcare leadership. *International Journal of Environmental Research and Public Health*, *19*(22), 15202. https://doi.org/10.3390/ijerph192215202

n8n.io. (2025). *n8n: Workflow automation tool*. GitHub. https://github.com/n8n-io/n8n

Nayebpour, H., & Sehhat, S. (2023). Designing the competency model of human resource managers based on paradox theory (Case study: Information and communication technology industry). *International Journal of Organizational Analysis*, *32*. https://doi.org/10.1108/IJOA-02-2023-3645

Nielsen Norman Group. (2025). *Dot voting: A simple decision-making and prioritizing technique in UX*. https://www.nngroup.com/articles/dot-voting/

OpenAI. (2024). *Swarm*. GitHub. https://github.com/openai/swarm

OpenAI. (2025). *OpenAI*. Ask ChatGPT Anything. https://openai.com/

Padzil, M. R., Karim, A., & Husnin, H. (2021). Employing DDR to design and develop a flipped classroom and project based learning module to applying design thinking in design and technology. *International Journal of Advanced Computer Science and Applications*, *12*. https://doi.org/10.14569/IJACSA.2021.0120988

Papakonstantinou, N., Van Bossuyt, D., Bell, R., Longshore, R., & Heikkila, M. (2025, January). *PrivateAIDELPHI: Adopting and adapting private AI for risk assessment of safety critical systems*. RAMS, Miramar Beach, FL. https://doi.org/10.1109/RAMS48127.2025.10935226

Priem, R. L., & Price, K. H. (1991). Process and outcome expectations for the dialectical inquiry, devil's advocacy, and consensus techniques of strategic decision making: *Group & Organization Management*, *16*(2), 206–225. https://doi.org/10.1177/105960119101600207

Punzi, C., Pellungrini, R., Setzu, M., Giannotti, F., & Pedreschi, D. (2024). AI, meet human: Learning paradigms for hybrid decision making systems. *arXiv.Org*, *abs/2402.06287*. https://doi.org/10.48550/arxiv.2402.06287

Rahman, N. H. N. A., & Kamauzaman, T. H. T. (2022). Developing key performance indicators for emergency department of teaching hospitals: A mixed fuzzy Delphi and nominal group technique approach. *Malaysian Journal of Medical Science*, *29*(2), 114–125. https://doi.org/10.21315/mjms2022.29.2.11

Rani, N. B. A., Hashim, M. E. A. B., Mustafa, W. A., Idris, M. Z. B., Al-Jawahry, H. M., & Ramadan, G. M. (2023). *Applying fuzzy Delphi method (FDM) to obtain the expert consensus in aesthetic experience (AX) and immersive experience (IX) elements for virtual reality historical event (VR historical event)*. 1–4. https://doi.org/10.1109/icmnwc60182.2023.10435812

Rogelberg, S. G., Barnes-Farrell, J. L., & Lowe, C. A. (1992). The stepladder technique: An alternative group structure facilitating effective group decision making. *Journal of Applied Psychology*, *77*(5), 730–737. https://doi.org/10.1037/0021-9010.77.5.730

Rogelberg, S. G., & O'Connor, M. S. (1998). Extending the stepladder technique: An examination of self-paced stepladder groups. *Group Dynamics: Theory, Research, and Practice*, *2*(2), 82–91. https://doi.org/10.1037//1089-2699.2.2.82

SciSpace. (2025). *SciSpace*. https://typeset.io/

Spranger, J., & Niederberger, M. (2025). How Delphi studies in the health sciences find consensus: A scoping review. *Systematic Reviews*, *14*(1). https://doi.org/10.1186/s13643-024-02738-3

SuperAGI. (2025). *SuperAGI: Autonomous AI agent framework*. GitHub. https://github.com/TransformerOptimus/SuperAGI

Tung, L. L., & Quaddus, M. (2001). Conflict management in dialectical inquiry, devil's advocacy and consensus-based decision making approaches in a GSS environment. *Proceedings of the Pacific Asia Conference on Information Systems (PACIS)*. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1094&context=pacis2001

Ullrika Sahlin (Director). (2023, February 7). *Structured expert judgment to assess uncertainty* [Video recording]. Lund University. https://www.youtube.com/watch?v=QPrawReZBxI

Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, *8*(12), 2293–2303. https://doi.org/10.1038/s41562-024-02024-1

Valerdi, R. (2005). *The constructive systems engineering cost model (COSYSMO)*.

Vedantham, S., Gloviczki, P., Carman, T. L., Schneider, O., Sabri, S. S., & Kolluri, R. (2023). Delphi consensus on reporting standards in clinical studies for endovascular treatment of acute iliofemoral venous thrombosis and chronic iliofemoral venous obstruction. *Circulation-Cardiovascular Interventions*, *16*, e012894. https://doi.org/10.1161/CIRCINTERVENTIONS.123.012894

Exploring Visual Question Answering Capabilities of Multi-Modal Large Language Models with Model Based Systems Engineering Models

Ryan Longshore—is a 20-year veteran of the defense and electric utility industries. At NIWC LANT, he leads teams developing and integrating new technologies into Navy command centers. He is involved in the Navy's digital engineering transformation with a focus on model-based systems and model-based engineering. Ryan holds a BS in electrical engineering (Clemson) and an MS in systems engineering (SMU), and is pursuing a PhD in systems engineering at the Naval Postgraduate School. He is a South Carolina Registered Professional Engineer and an INCOSE CSEP, and holds the OMG SysML Model Builder Fundamental Certification. [ryan.longshore@nps.edu]

Ryan Bell—is an 9-year experienced engineer in the defense industry. In his current role at Naval Information Warfare Center Atlantic (NIWC LANT), Ryan 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 life cycle. Ryan earned a BS in electrical engineering from Clemson University and a MS in electrical engineering from Clemson University with a focus on electronics, and 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.bell@nps.edu]

Raymond Madachy, PhD—is a Professor in the Systems Engineering Department at the Naval Postgraduate School. His research interests include systems engineering tool environments for digital engineering, modeling and simulation of systems and software engineering processes, generative AI, and system cost modeling. He has developed cost estimation tools for systems and software engineering, and created the Systems Engineering Library (se-lib). His books include Software Process Dynamics, What Every Engineer Should Know about Modeling and Simulation, What Every Engineer Should Know about Python, and he co-authored Software Cost Estimation with COCOMO II and Software Cost Estimation Metrics Manual for Defense Systems. [rjmadach@nps.edu]

Abstract

The continued advancement of large language models (LLMs) has unlocked new opportunities for systems engineering particularly in the field of visual question answering (VQA). Multi-modal LLMs are capable of processing both textual and graphical inputs, allowing them to interpret the graphical elements of model-based systems engineering (MBSE) models alongside accompanying textual descriptions. This paper explores the capabilities of multi-modal LLMs in understanding and interpreting Systems Modeling Language (SysML) v1 block definition diagrams (BDDs). BDDs are visual diagrams that formally capture a system's structural elements, properties, relationships, and multiplicities.

We evaluate both proprietary and open-source multi-modal LLMs using a curated dataset of SysML BDDs and associated multiple-choice question set designed to assess LLM performance at the first two levels of Bloom's Taxonomy, Remember and Understand. We also analyzed the effect of model size on accuracy. The results provide insights into which current LLMs are able to natively interpret SysML BDD syntax which informs future research aimed at enhancing systems modeling processes with AI agents.

Introduction

The integration of artificial intelligence (AI) into Model-Based Systems Engineering (MBSE) processes presents significant opportunities for improving model comprehension, validation, and support activities. Multi-modal large language models (LLMs) are capable of processing both textual and graphical inputs and have expanded the potential for automating the interpretation of system modeling language (SysML) v1 models. Block Definition Diagrams

(BDDs) are key elements of SysML v1 models, serving as a foundational representation of system structure, properties, and relationships (OMG, 2019).

Despite the rapid evolution of LLMs, their ability to accurately interpret SysML artifacts remains largely unexplored. Existing evaluations of multi-modal LLMs have primarily focused on images or general diagrammatic reasoning, rather than domain-specific graphical languages such as SysML (Antol et al., 2015; Ishmam et al., 2024; Lin et al., 2014). This gap limits the current understanding of LLMs' effectiveness in supporting engineering workflows that rely on formal SysML model interpretation.

This paper addresses this gap by evaluating the performance of contemporary multi-modal LLMs in interpreting SysML v1.x BDDs. We develop a curated dataset of BDDs and design a multiple-choice question set aligned with the first two levels of Bloom's Taxonomy. The evaluation examines both proprietary and open-source LLMs, analyzing their capabilities across models of varying sizes. The findings offer empirical insights into the strengths and limitations of current LLMs in understanding formal systems modeling artifacts and inform future research on enhancing AI-driven support for MBSE practices.

Background and Related Research

Visual Question Answering

Visual question answering (VQA) is a field of AI research focused on answering textual questions using image(s) as contextual input (Antol et al., 2015). Responses can be binary (yes/no), multiple choice, or open-ended. Early VQA methods combined computer vision (CV) feature extraction and natural language processing (NLP) machine learning (ML) techniques to generate answers (Ishmam et al., 2024). The introduction of attention mechanisms such as stacked attention networks and dynamic memory networks enabled multi-step reasoning in VQA tasks (Xiong et al., 2016; Yang et al., 2016). Large Visual Language Models (LVLMs) such as ViLBERT and VisualBERT further advanced the field by incorporating pretraining techniques and transformer architecture to increase model performance (Li et al., 2019; Lu et al., 2019).

The emergence of multi-modal LLMs transformed VQA by enabling unified reasoning over text and images. Models like Flamingo and PaLI demonstrated that scaling vision-language pretraining yields strong few-shot VQA capabilities (Alayrac et al., 2022; Chen et al., 2023). BLIP-2 (Li et al., 2023) further streamlined this approach by efficiently connecting frozen pre-trained image encoders and LLMs (Li et al., 2023). OpenAI's GPT-4 represented a shift toward general-purpose multi-modal reasoning achieving similar performance to text only inputs without VQA-specific architectures (OpenAI et al., 2024). These advancements have moved VQA from specialized models toward foundation models with broad applicability across engineering and scientific tasks.

VQA Benchmarks

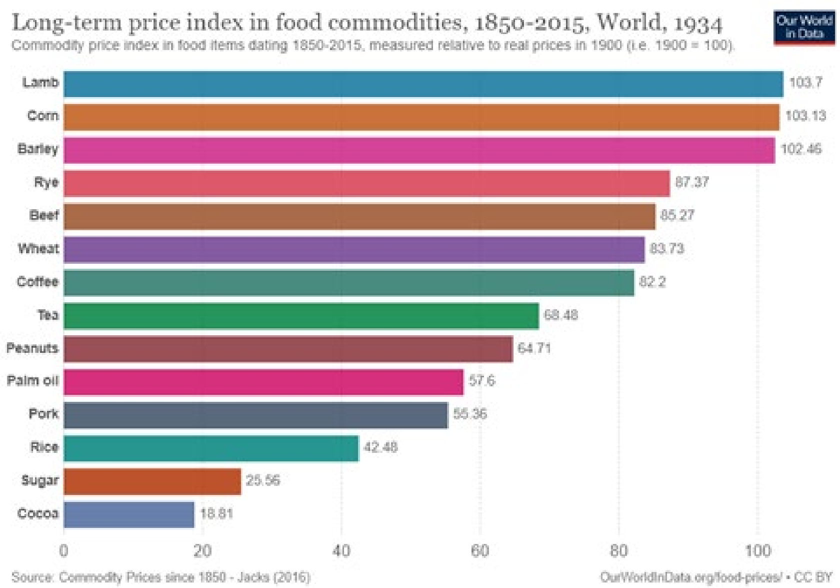
A variety of datasets have been developed to benchmark VQA capabilities. The Dataset for Question Answering on Real-Work images (DAQUAR) was one of the first largely used VQA benchmarks and was a modern attempt at a "visual Turing test" (Malinowski & Fritz, 2015). Microsoft's Common Objects in Context (COCO) dataset introduced a large dataset where each image was provided as a raw image and then a segmented image with highlighted objects (Figure 1) that enabled benchmarking for tasks such as counting (Lin et al., 2014). The VQA-2.0 dataset balanced the VQA-1.0 dataset by collecting complementary images for each question ensuring that each question could be applied to different images and yield different answers (Antol et al., 2015; Goyal et al., 2019).

(a) (b)

Figure 9. Sample Image from COCO Dataset (*COCO - Common Objects in Context*, n.d.) (a) Original Image, (b) Segmented Image Displaying Overlay for Fire Hydrants and Vehicles

While datasets such as DAQUAR, COCO, and VQA 2.0 addressed general VQA questions, they also highlighted the need for application specific datasets such as chart and diagram specific datasets. These chart and diagram specific datasets sought to address recalling and synthesizing data from the chart using methods such as optical character recognition (OCR), interpreting numerical data contained in a chart, and understanding of different chart structures (Kafle et al., 2020).

Chart and figure specific datasets continue to evolve along with new methods to improve complex reasoning (Srivastava & Sharma, 2024). Data visualization question answering (DVQA) is one of the early chart datasets specifically focused on bar charts (Kafle et al., 2018). When introducing the DVQA dataset, Kafle et al. showed that VQA methods were not effective at recalling or synthesizing data related to bar charts and proposed new methods for chart specific VQA. Also introduced in 2018, FigureQA expanded charts types to include line plots, dot-line plots, and pie charts in addition to bar charts and proposed a Relation Network method (Kahou et al., 2018). More recent datasets such as ChartQA introduce complex reasoning questions that require logical and arithmetic calculations (Masry et al., 2022). A sample from the ChartQA dataset shown in Figure 2 demonstrates the increased complexity of questions. Answering the questions requires the number of bars in the chart, analyzing their labels for relevance (is it a food or not), and then combining those two pieces of information to determine the correct answer.



Question 1 in the ChartQA 'test' dataset:

Q: How many food items are shown in the bar graph?
A: 14

Figure 10. ChartQA Sample Question and Associated Image (*Lmms-Lab/ChartQA · Datasets at Hugging Face*, 2024)

SysML v1.6 BDDs

In SysML v1.6, "the BDD is used to define blocks in terms of their features, and their structural relationships with other blocks" (Friedenthal et al., 2011). While a BDD can convey many types of information about blocks and their relationships, this paper focuses on the following parts of the BDD as described in *SysML Distilled* (Delligatti, 2014):

- Blocks are fundamental modeling elements that represent system components, subsystems, or other concepts (e.g., actors). They can define both structural and behavioral features.
- Properties are attributes owned by a block that define the internal structure and characteristics.
    - Part properties represent a block's internal structure and are used to model composition.
    - Reference properties represent a relationship to an external structure and are used to show dependency on another block.
    - Value properties represent a quantitative or descriptive attribute of a block (e.g., speed in miles per hour, length in inches)
- Relationships convey composition, abstraction, connection, or dependencies between model elements.
    - Composite associations convey structural decomposition and are denoted by filled in diamonds.

- Reference associations convey a connection or dependency between two blocks. They may also be shown as reference properties.
- Generalizations convey inheritance between elements and are denoted by unfilled triangles. The generalized element is known as the supertype while the more specialized element is known as the subtype.

- Multiplicity is a constraint specifying the number of allowable instances, such as one-to-one (1) and one-to-many (1..*). Multiplicity can also be used to model optional components (0..1, 0..*).

Representing the LLM Cognitive Process with Bloom's Revised Taxonomy

Bloom's taxonomy is a hierarchical model of cognition widely used in education to classify learning objectives (Bloom et al., 1956). Bloom's revised taxonomy specifies six cognitive process levels: Remember, Understand, Apply, Analyze, Evaluate, and Create (Krathwohl, 2002). In addition to human cognition, recent research has extended Bloom's revised taxonomy to LLMs.

A recent study analyzing the alignment of existing LLM benchmarks to Bloom's revised taxonomy found that most benchmarks adequately assess the "Remember" and "Understand" levels but do not comprehensively address all six cognitive levels (Huber & Niklaus, 2025). Although "Remember" and "Understand" represent the lowest levels of cognition, LLMs do not always perform the highest at these levels. In a mixed-methods study examining ChatGPT's performance on psychosomatic medicine examination questions, researchers observed that GPT-4 exhibited notable deficiencies in these two levels, with 29 errors in "Remember" and 23 errors in "Understand" stemming from difficulties in recalling specific details and grasping conceptual relationships (Herrmann-Werner et al., 2024).

Consistent with other evaluation approaches, this study focuses on the first two levels of Bloom's revised taxonomy: Remember and Understand. "Remember" questions are designed to assess recall of information directly from SysML BDDs without requiring synthesis of multiple elements. "Understand" questions assess higher cognitive engagement through summarization and inference tasks. Summarization questions require synthesis of multiple pieces of information from the diagram while inference questions involve drawing conclusions that are not explicitly stated but are logically supported by the diagram's structure and consistent with SysML v1.6 rules.

Methodology

This section describes the methodology for constructing and evaluating a dataset aimed at assessing LLMs' ability to interpret SysML v1.x BDDs. In the absence of existing datasets focused specifically on SysML, a novel dataset was developed to capture both syntactic and semantic understanding of BDDs through structured multiple-choice questions aligned with Bloom's revised taxonomy (Krathwohl, 2002). A set of both proprietary and open source multi-modal LLMs were evaluated against this dataset. LLM inference was conducted using GPU-accelerated environments and automated through scripting to ensure consistency and reproducibility. The evaluation process culminated in a human as judge assessment of LLM responses where the human judge was a practicing systems engineer.

Dataset Generation

While there are several datasets focused on VQA and diagrams in particular, there are no datasets specifically focused on SysML v1.x. Therefore, the dataset for this analysis was generated by systems engineers with experience in both systems modeling and benchmark dataset generation. It consists of a curated set of SysML BDDs and associated multiple choice questions. The dataset was exclusively human-generated with no synthetic content.

The dataset consists of 80 questions. Generated questions cover four key concepts from SysML v1.x BDDs: Blocks, Properties, Relationships, and Multiplicity. The difficulty of the generated questions is evenly distributed across the remember and understand levels of Bloom's revised taxonomy. Table 1 details the distribution of questions across both Bloom's Taxonomy and BDD concept.

Table 1. Distribution of Questions

	Blocks	Properties	Relationships	Multiplicity
Remember	10	10	10	10
Understand	10	10	10	10

The dataset follows a syntax common to multiple choice question datasets with some minor modifications to incorporate additional fields such as diagram reference, SysML Concept, and Bloom Taxonomy Category as shown in Table 2. This syntax will allow the dataset to be easily expanded to more diagram types and potentially be incorporated as an extension into systems engineering specific benchmarks such as SysEngBench (Bell, 2024).

Table 2. Dataset Fields

Field	Data Format	Description
QuestionID	Integer	Unique identifier for each question
BDDConcept	Enumeration	One of four options: Blocks, Relationships, Properties, Multiplicity
BloomCategory	Enumeration	One of two options: Remember, Understand
Diagram	String	File name of the associated SysML BDD
Question	String	Text of the multiple choice question
ChoiceA	String	Text for choice A
ChoiceB	String	Text for choice B
ChoiceC	String	Text for choice C
ChoiceD	String	Text for choice D
Answer	String	Correct Answer: ChoiceA, ChoiceB, ChoiceC, ChoiceD

A camera specification BDD from the dataset is shown in Figure 3. This diagram incorporates blocks, value properties, a generalization relationship (denoted by the unfilled triangle), and other elements such as requirements and value types. Two sample questions based on this diagram are shown below. Note that the rationale field is included as a courtesy explanation to the reader as to why the answer is correct, but is not included in the dataset.

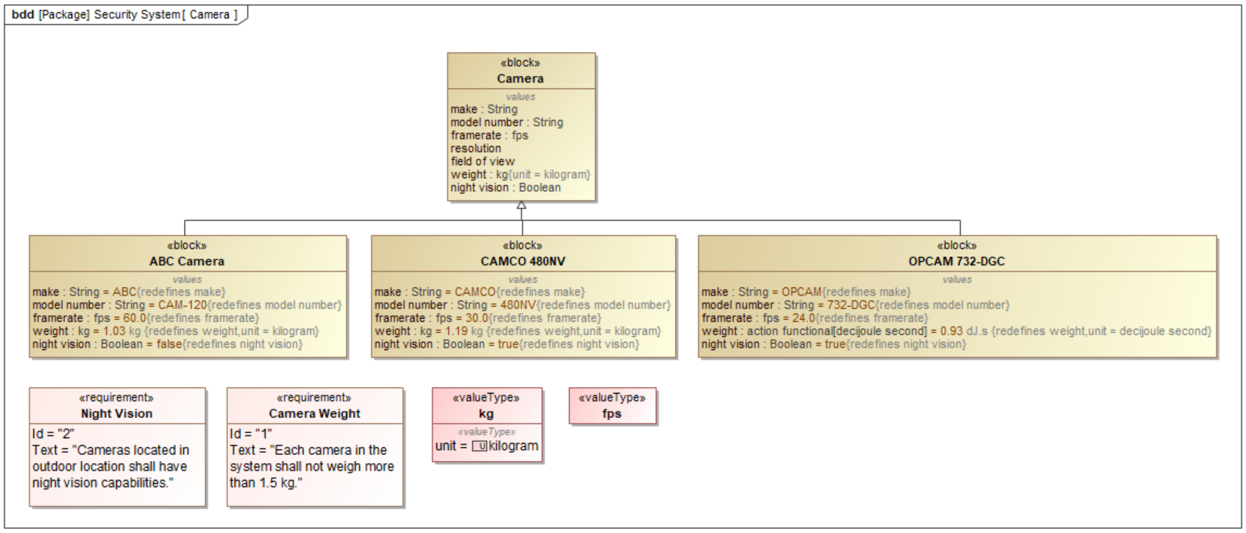


Figure 3. Camera Specification Diagram

A sample “Remember” question related to the BDD properties concept is:

What is the custom value type defined for framerate?

- *a) string*
- *b) kg*
- *c) boolean*
- *d) fps*

Correct Answer: d

Rationale: *The framerate property is typed by the fps property in the ‘framerate : fps’ value property definition.*

A sample *understand* question related to the BDD properties category is:

How many value properties are there for each camera?

- *a) 15*
- *b) 22*
- *c) 5*
- *d) 7*

Correct Answer: d

Rationale: *The generalized camera block contains seven value properties that are inherited by each camera. Each specific camera block shows the five value properties that are re-defined, but not the inherited properties that are not re-defined.*

LLM Selection

A variety of open source and proprietary models easily accessible to practicing engineers were selected for this paper. The open source models were selected as they are the multi-modal vision models available from the widely used Ollama library as of April 2025

(*Ollama*, n.d.). ChatGPT-4o and Sonnet-3.7 were selected as they are widely available proprietary models. The dataset was evaluated against the following models:

- baklava: 7B
- gemma3: 4B, 12B, and 27B Variants
- granite3.2-vision: 2B
- llama3.2-vision: 11B and 90B Variants
- llava: 7B, 13B and 34B Variants
- llava-llama3: 8B
- minicpm-v: 8B
- mistral-small3.1: 24B
- moondream: 1.8B
- OpenAI chatgpt-4o
- Anthropic sonnet-3.7

Dataset Evaluation

To evaluate the dataset using Ollama models, a virtual GPU pod instance was provisioned on RunPod utilizing an NVIDIA A40 GPU. Ollama was installed on this virtual pod following the guidelines provided in the RunPod documentation (*Set up Ollama on Your GPU Pod | RunPod Documentation*, 2025). The selected LLMs were then loaded into the GPU pod via the ollama *run* and *pull* commands. A Jupyter Notebook was deployed within the same pod to facilitate the evaluation process. The question set formatted as a CSV file along with the corresponding images was uploaded to the notebook environment. A Python script was developed to automate the process of asking questions to the LLMs and capturing their responses. The script iterated through each question in the question set, submitted each prompt to the LLM under evaluation, and recorded the generated answers. The outputs were then written to a CSV file for analysis. This workflow is detailed in Figure 4.

The same dataset was used to evaluate the chatgpt-4o and sonnet-3.7 models. However, instead of using custom scripts and dedicated GPUs, the ChatGPT (*ChatGPT*, n.d.) and Claude (*Claude*, n.d.) websites were utilized to ask the LLMs questions.

LLMs do not explicitly know they should answer a multiple choice question with a one character response. Therefore the question was asked in the following format:

You are an automated system that answer multiple choice questions and only outputs one of four letters: A, B, C, or D. Given the following question and four answer choices, respond with ONLY the letter of the best answer. This will be A, B, C, or D. Do not explain your answer. Do not say anything else. Use the image as context for your answer.

Question: {question}

A. {option_a}

B. {option_b}

C. {option_c}

D. {option_d}

There are several methods to compare the model answers to the correct answers including LLM as a judge and human as a judge. LLM as a judge refers to the use of LLMs as automated judges for evaluating other LLMs on open-ended tasks where traditional benchmarks may be insufficient (Zheng et al., 2023). However, due to limitations in dataset size (80 questions), model coverage (18 models), and the fact that LLM judging focuses on evaluating the final answer rather than the reasoning process behind it, human as a judge is employed for the final assessment.

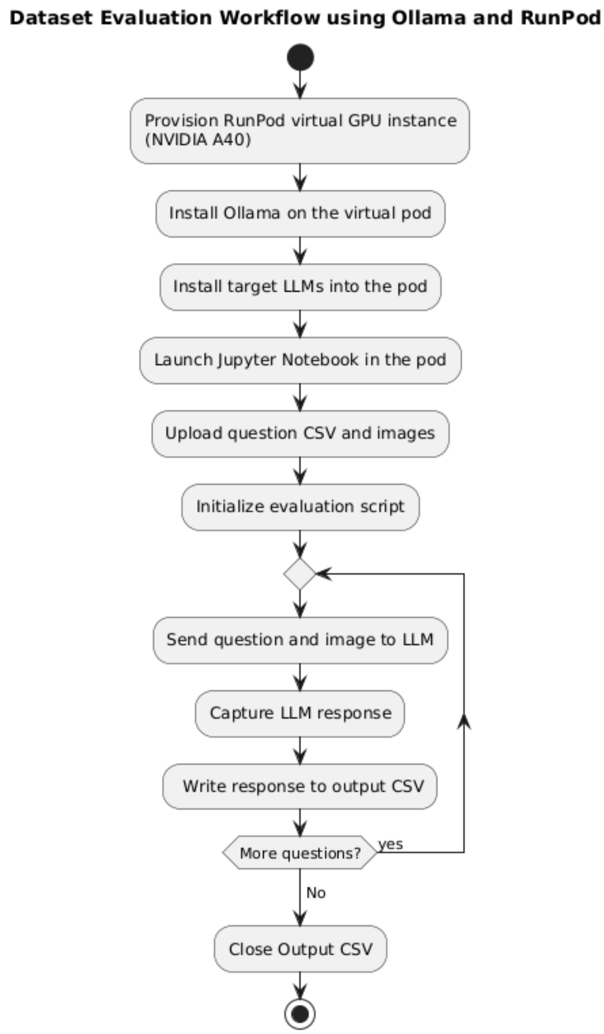


Figure 4. Dataset Evaluation Workflow

Results, Discussion, and Limitations

Each LLM's responses were scored against the correct multiple-choice answers to evaluate accuracy. Accuracy is defined as the percentage of questions answered correctly. The dataset was designed to assess both syntactic and semantic understanding of SysML BDDs covering a balanced distribution across four modeling concepts and two levels of Bloom's revised taxonomy. The results presented below compare overall model performance, analyze trends relative to model size, and break down accuracy by cognitive level and SysML concept.

The overall performance of each LLM is shown in Figure 5. Proprietary LLMs are denoted by orange bars while open source LLMs are denoted by blue bars. Although proprietary models secured two of the top three scores, the open-source model mistral-small3.1, a 24B model, outperformed Sonnet-3.7 while falling short of GPT-4o. Given that each multiple-choice question included four possible answers, the expected accuracy from random guessing across all 80 questions is 25%. Bakllava, a 7B model, demonstrated the lowest performance and was the only model that failed to exceed the random guessing baseline.

The scatter plot in Figure 6 compares LLM accuracy to model size. It is important to note that the size of GPT-4o and Sonnet-3.7 is not publicly available information. There are several estimates of around 200 billion parameters, but those estimates have not been confirmed by either OpenAI or Anthropic. A correlation coefficient of 0.65 indicates a moderate relationship between LLM size and accuracy. However, mistral-small3.1 (24B) outperforms three larger open source models as well as Sonnet-3.7. Despite being the second smallest model, granite3.2-vision (2B) outperforms 10 larger models. These observations suggest that factors beyond parameter count, such as training data and/or methods, influence performance.

The grouped bar chart in Figure 7 visualizes accuracy by Bloom's revised taxonomy category. Most LLMs perform better on "Remember" tasks than on "Understand" questions with GPT-4o correctly answering all "Remember" questions. Two LLMs performed slightly better on "Understand" questions. These results indicate the LLMs' ability to recall information from a diagram is greater than the ability to synthesize multiple pieces of information or bring in additional context not explicitly stated in the BDD.

The multi-series bar chart shown in Figure 8 breaks down performance across the four core SysML v1.6 BDD concepts: Blocks, Relationships, Properties, and Multiplicity. The results reveal notable variation across concepts, with most models performing best on Blocks and worst on Relationships or Multiplicity, highlighting uneven conceptual understanding amongst LLMs.

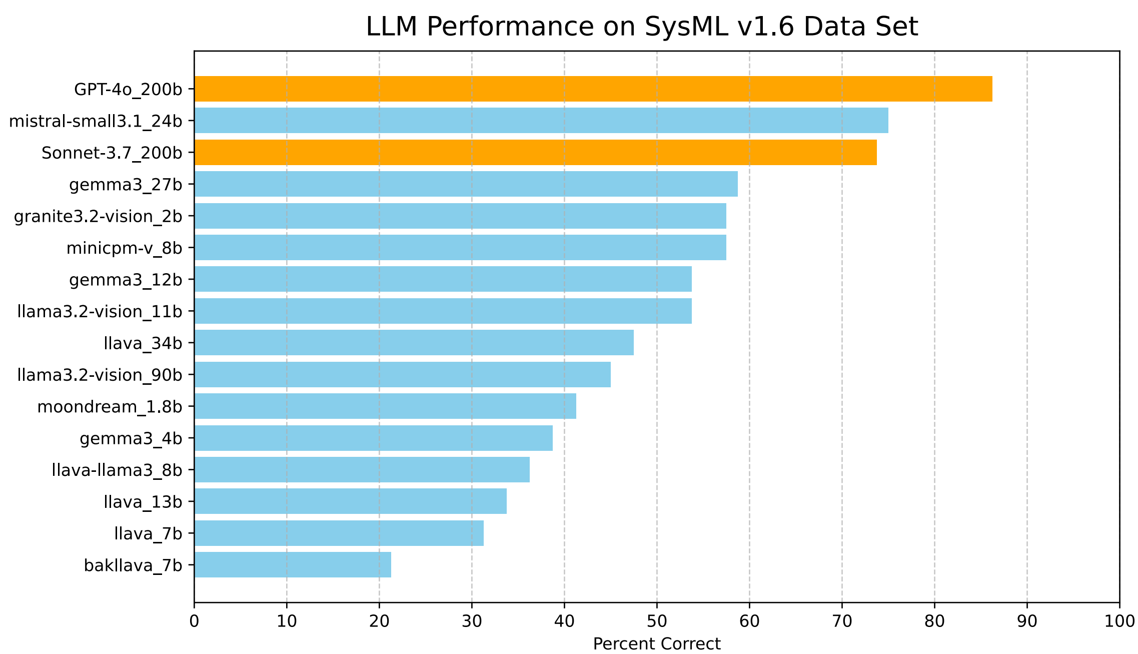


Figure 5. Overall Performance

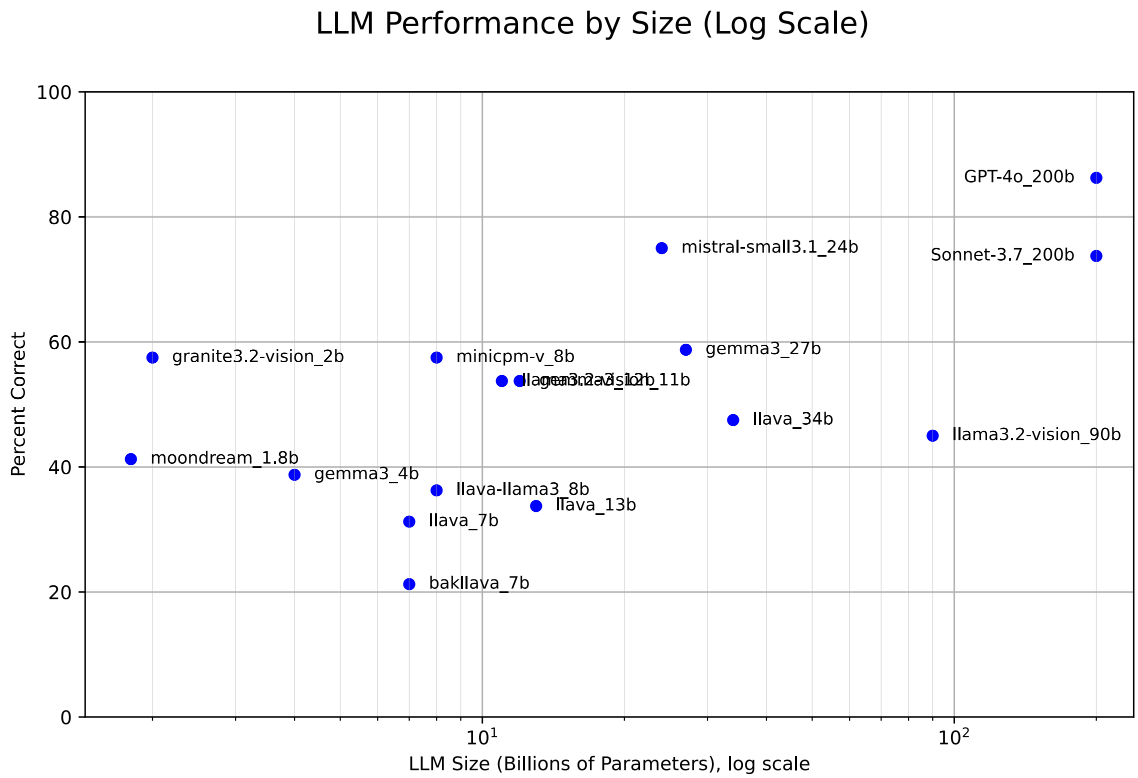


Figure 6. Performance by LLM Size

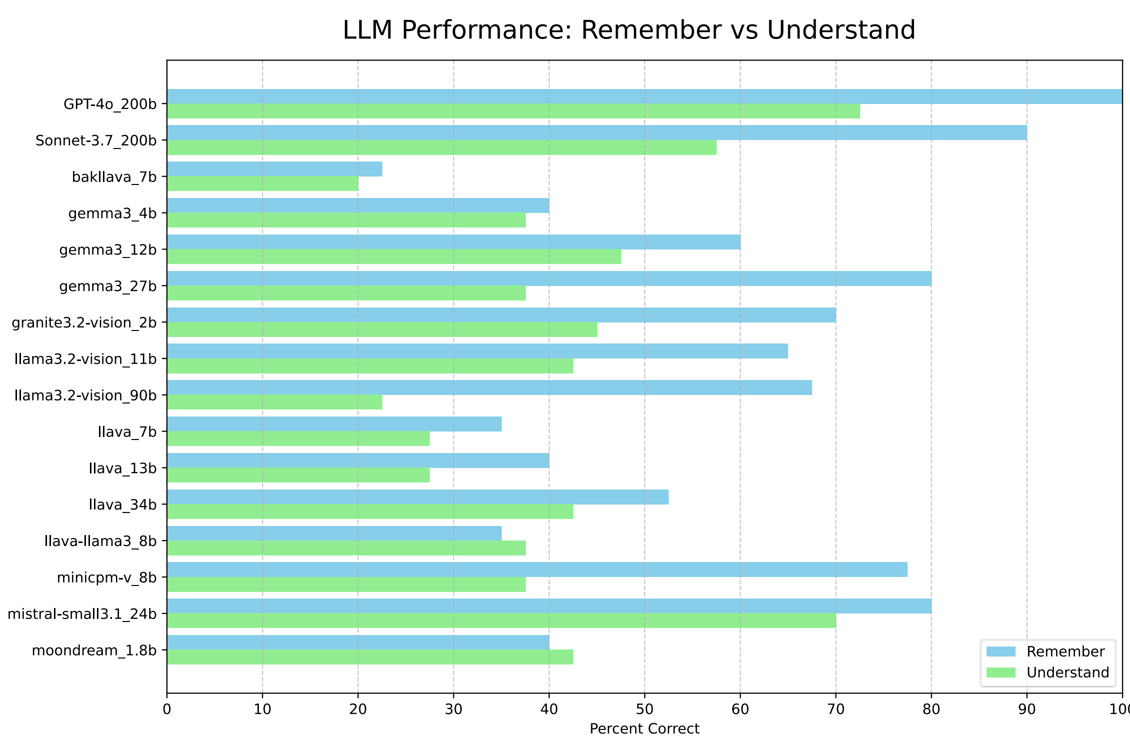


Figure 7. Performance by Bloom's Taxonomy Category

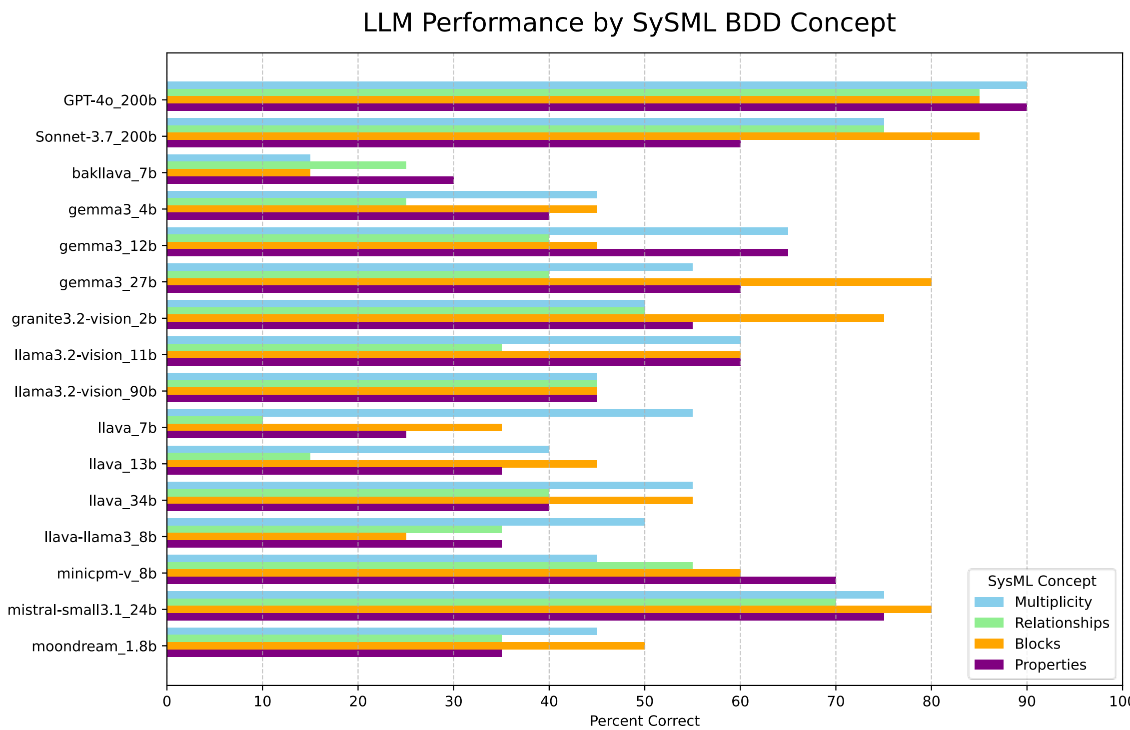


Figure 8. Performance by BDD Concept

Future Work

This study is an initial exploration of LLM performance on SysML v1.x BDDs using a curated dataset of multiple-choice questions. While the current dataset is balanced across four core BDD concepts and two levels of Bloom's revised taxonomy ("Remember" and "Understand"), future research can extend the depth and breadth of this analysis.

Future studies could focus on expanding the dataset in three different ways:

- Expansion of the dataset to include more questions and images. This could increase the robustness of the evaluation and potentially increase the statistical significance of the results.
- Incorporation of additional SysML v1.x diagram types beyond BDDs such as Internal Block Diagrams (IBDs), Activity Diagrams, and Sequence Diagrams would provide a more comprehensive benchmark to evaluate the extent to which LLMs can generalize across different visual and semantic structures in systems modeling. This would also increase the number of multiple-choice questions per Bloom's revised taxonomy level to improve statistical robustness and reduce sensitivity to specific wording or diagram features.
- Expansion of the dataset to include higher levels of Bloom's revised taxonomy, such as "Apply," "Analyze," "Evaluate," and potentially even "Create" could give a more holistic view of LLM capabilities. By incorporating these more complex cognitive tasks, future studies can investigate whether LLM performance declines as tasks become more abstract and cognitively demanding.

This study identified several LLMs that may be promising candidates for techniques such as Retrieval-Augmented Generation (RAG) to improve accuracy. Applying RAG could allow models to draw from relevant SysML documentation or design patterns to enhance their question answering abilities. Future experiments could explore the impact of RAG on accuracy particularly in handling the more difficult “Understand” questions or tasks at higher levels or Bloom’s revised taxonomy.

Conclusion

This study presents a targeted evaluation of multi-modal LLMs on SysML v1.6 BDDs through a VQA framework. By grounding the analysis in Bloom’s revised taxonomy and assessing both proprietary and open-source models, we provide empirical insights into how LLMs interpret formal, domain-specific systems modeling diagrams. The findings show that while model size moderately correlates with accuracy, other factors also impact LLM performance. Most models demonstrate stronger capabilities in recalling elements (“Remember”) than in synthesizing or inferring information (“Understand”) revealing limitations in semantic comprehension of structured graphical artifacts.

The curated dataset and evaluation framework introduced here lay the groundwork for future research into more advanced cognitive tasks and broader SysML diagram types. As the field progresses, improving model performance through techniques like RAG on domain-specific content holds significant promise. Ultimately, understanding and enhancing how LLMs process systems modeling artifacts is a critical step toward their meaningful integration into MBSE workflows.

Acknowledgements

This work has benefited from the use of generative AI tools including ChatGPT and SciSpace for writing assistance and code development. These tools were employed to enhance conceptual clarity, improve code efficiency, and support literature synthesis. All AI-generated outputs were critically reviewed, refined, and validated to ensure accuracy and alignment with academic integrity. Their contributions were limited to supporting the research process, and final responsibility for the content, analysis, and conclusions remains with the authors.

References

Alayrac, J.-B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., Lenc, K., Mensch, A., Millican, K., Reynolds, M., Ring, R., Rutherford, E., Cabi, S., Han, T., Gong, Z., Samangooei, S., Monteiro, M., Menick, J., Borgeaud, S., … Simonyan, K. (2022). *Flamingo: A visual language model for few-shot learning* (No. arXiv:2204.14198). ArXiv. https://doi.org/10.48550/arXiv.2204.14198

Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C. L., & Parikh, D. (2015). VQA: Visual question answering. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2425–2433. https://doi.org/10.1109/ICCV.2015.279

Bell, R. (2024, August). Rabell/SysEngBench \cdot datasets at Hugging Face. In *SysEngBench*. https://huggingface.co/datasets/rabell/SysEngBench

Bloom, B. S., Engelhart, M. D., Furst, E., Hill, W. H., & Krathwohl, D. R. (1956). *Handbook I: Cognitive domain*. David McKay, 483–498.

ChatGPT. (n.d.). Retrieved April 24, 2025, from https://chatgpt.com

Chen, X., Wang, X., Changpinyo, S., Piergiovanni, A. J., Padlewski, P., Salz, D., Goodman, S., Grycner, A., Mustafa, B., Beyer, L., Kolesnikov, A., Puigcerver, J., Ding, N., Rong, K., Akbari, H., Mishra, G., Xue, L., Thapliyal, A., Bradbury, J., … Soricut, R. (2023). *PaLI: A jointly-scaled multilingual language-image model* (No. arXiv:2209.06794). ArXiv. https://doi.org/10.48550/arXiv.2209.06794

Claude. (n.d.). Retrieved April 24, 2025, from https://claude.ai/new

COCO - Common Objects in Context. (n.d.). Retrieved April 20, 2025, from https://cocodataset.org/#explore

Delligatti, L. (2014). *SysML distilled*. Pearson Education.

Friedenthal, S., Moore, A., & Steiner, R. (2011). *A practical guide to SysML: The systems modeling language*. Elsevier Science & Technology. http://ebookcentral.proquest.com/lib/ebook-nps/detail.action?docID=787244

Goyal, Y., Khot, T., Agrawal, A., Summers-Stay, D., Batra, D., & Parikh, D. (2019). Making the V in VQA matter: Elevating the role of image understanding in visual question answering. *International Journal of Computer Vision*, *127*(4), 398–414. https://doi.org/10.1007/s11263-018-1116-0

Herrmann-Werner, A., Festl-Wietek, T., Holderried, F., Herschbach, L., Griewatz, J., Masters, K., Zipfel, S., & Mahling, M. (2024). Assessing ChatGPT's mastery of Bloom's taxonomy using psychosomatic medicine exam questions: Mixed-methods study. *Journal of Medical Internet Research*, *26*, e52113. https://doi.org/10.2196/52113

Huber, T., & Niklaus, C. (2025). LLMs meet Bloom's taxonomy: A cognitive view on large language model evaluations. In O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, & S. Schockaert (Eds.), *Proceedings of the 31st International Conference on Computational Linguistics* (pp. 5211–5246). Association for Computational Linguistics. https://aclanthology.org/2025.coling-main.350/

Ishmam, Md. F., Shovon, Md. S. H., Mridha, M. F., & Dey, N. (2024). From image to language: A critical analysis of visual question answering (VQA) approaches, challenges, and opportunities. *Information Fusion*, *106*, 102270. https://doi.org/10.1016/j.inffus.2024.102270

Kafle, K., Price, B., Cohen, S., & Kanan, C. (2018). DVQA: Understanding data visualizations via question answering. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5648–5656. https://doi.org/10.1109/CVPR.2018.00592

Kafle, K., Shrestha, R., Price, B., Cohen, S., & Kanan, C. (2020). Answering questions about data visualizations using efficient bimodal fusion. *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1487–1496. https://doi.org/10.1109/WACV45572.2020.9093494

Kahou, S. E., Michalski, V., Atkinson, A., Kadar, A., Trischler, A., & Bengio, Y. (2018). *FigureQA: An annotated figure dataset for visual reasoning* (No. arXiv:1710.07300). ArXiv. https://doi.org/10.48550/arXiv.1710.07300

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, *41*(4), 212.

Li, J., Li, D., Savarese, S., & Hoi, S. (2023). *BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models* (No. arXiv:2301.12597). ArXiv. https://doi.org/10.48550/arXiv.2301.12597

Li, L. H., Yatskar, M., Yin, D., Hsieh, C.-J., & Chang, K.-W. (2019). *VisualBERT: A simple and performant baseline for vision and language* (No. arXiv:1908.03557). arXiv. https://doi.org/10.48550/arXiv.1908.03557

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In D. Fleet, T. Pajdla, B. Schiele, &

T. Tuytelaars (Eds.), *Computer vision—ECCV 2014* (pp. 740–755). Springer International Publishing. https://doi.org/10.1007/978-3-319-10602-1_48

Lmms-lab/ChartQA · Datasets at Hugging Face. (2024, October 4). https://huggingface.co/datasets/lmms-lab/ChartQA

Lu, J., Batra, D., Parikh, D., & Lee, S. (2019). *ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language rasks* (No. arXiv:1908.02265). ArXiv. https://doi.org/10.48550/arXiv.1908.02265

Malinowski, M., & Fritz, M. (2015). *A multi-world approach to question answering about real-world scenes based on uncertain input* (No. arXiv:1410.0210). ArXiv. https://doi.org/10.48550/arXiv.1410.0210

Masry, A., Do, X. L., Tan, J. Q., Joty, S., & Hoque, E. (2022). ChartQA: A benchmark for question answering about charts with visual and logical reasoning. In S. Muresan, P. Nakov, & A. Villavicencio (Eds.), *Findings of the association for computational linguistics: ACL 2022* (pp. 2263–2279). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.findings-acl.177

Ollama. (n.d.). *Vision models · Ollama Search*. Retrieved April 18, 2025, from https://ollama.com/search

OMG. (2019, November). *OMG Systems Modeling Language version 1.6*. https://www.omg.org/spec/SysML/1.6/PDF

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2024). *GPT-4 technical report* (No. arXiv:2303.08774). ArXiv. https://doi.org/10.48550/arXiv.2303.08774

Set up Ollama on your GPU Pod | RunPod Documentation. (2025). https://docs.runpod.io/tutorials/pods/run-ollama

Srivastava, S., & Sharma, H. (2024). Deep multimodal relational reasoning and guided attention for chart question answering. *Journal of Electronic Imaging*, *33*(6), 063052. https://doi.org/10.1117/1.JEI.33.6.063052

Xiong, C., Merity, S., & Socher, R. (2016). *Dynamic memory networks for visual and textual question answering* (No. arXiv:1603.01417). ArXiv. https://doi.org/10.48550/arXiv.1603.01417

Yang, Z., He, X., Gao, J., Deng, L., & Smola, A. (2016). *Stacked attention networks for image question answering* (No. arXiv:1511.02274). ArXiv. https://doi.org/10.48550/arXiv.1511.02274

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E. P., Zhang, H., Gonzalez, J. E., & Stoica, I. (2023). *Judging LLM-as-a-judge with MT-Bench and chatbot arena* (No. arXiv:2306.05685). ArXiv. https://doi.org/10.48550/arXiv.2306.05685

THIS PAGE INTENTIONALLY LEFT BLANK

Identifying Downtime Drivers Using SIMLOX Simulations to Rapidly Develop Solutions Improving System and Mission Readiness

Marissa McLaren—is an Analyst at Systecon North America working on lifecycle management and sustainment for complex technical systems. Utilizing Systecon's proprietary software for optimization and simulation modeling, she has worked with PEO IWS, NAVSUP, NAVSEA, NAVAIR, and JPO. She has an MS in Engineering and Public Policy from Carnegie Mellon University, and a BS in Mechanical Engineering from Georgia Institute of Technology. [marissa.mclaren@systecon.us]

Abstract

In this paper we present a simulation model that directly relates the system results to the individual parts and resources, otherwise known as downtime drivers. Understanding operational downtime drivers and whether they are item- or subsystem-specific is crucial to making strategic decisions for missions and operations. Identifying the drivers in the modeling phase allows for increased preparation and problem-solving to improve mission requirements. While working to solve the issues created by downtime drivers, industry and defense can work together to determine a reasonable solution to overcome the impact an item or subsystem can have on the overall system.

This case study describes a scenario where industry and defense have been able to identify downtime drivers for a complicated system and develop a set of reasonable alternatives to address these issues. OPUS10 identifies the initial spares purchase optimization for a given availability requirement. We can then utilize that recommendation in SIMLOX, a Monte Carlo–based simulation tool. Simulation results are often used to identify bottlenecks within the supply chain, spares, and support organization. With the recent software updates to SIMLOX, we can identify downtime drivers. Stakeholders can identify which subsystem(s) or items are causing a system to be down.

Keywords: Simulation, Downtime Drivers, System Readiness, Solution-Focused Modeling

Background

Systecon North America provides software support, training, and consulting for their proprietary software, OPUS Suite. OPUS Suite comprises three main software solutions. These include OPUS (optimization tool), SIMLOX (simulation tool) and CATLOC (cost control tool). Systecon has also developed EVO (tactical operation), INSIGHTS (business intelligence), and CONNECT (integration). Having a basic understanding of OPUS and SIMLOX will be beneficial to understanding the benefit of rapid downtime driver identification.

OPUS is a tool used for cost-effective spare parts steady-state optimization, balancing the spare part investment while also maintaining or increasing system readiness. Beyond just spares optimization, OPUS allows the user to evaluate support solutions, technical systems, and scenarios (Systecon Group, 2020*)*. OPUS is utilized across the Department of Defense as a tool for spares part purchases, such as the Navy Common Readiness Model, as well as for stress testing what-if scenarios for large programs such as the F-35 Program (Systecon North America*,* 2019).

SIMLOX is an event-driven simulation tool that allows detailed analysis of technical systems' performance over time, while factoring in varying operational and logistics support scenarios (Systecon Group*,* 2018). Since SIMLOX includes the time component, you can identify weaknesses in the support structure during a peak utilization period. The flexibility allows one to model any complex technical system. For example, their varied usage and utilization, varied support capabilities over time, complex functional block diagrams, performance targets, confidence intervals, and, most important for this paper, downtime drivers.

Introduction

Complex defense systems with intricate missions rely on cost optimization and simulation models to determine the best set of variables to meet mission and system readiness requirements. Linking system properties and logistics support to understand system performance is a multidimensional optimization problem over multi-indenture systems that requires an iterative approach. In a basic model, one could factor each variable to understand the overall system performance. However, for complex systems and support structures, an analytic solution becomes untenable. One cannot look at each item, system, mission, and resource in isolation; they all have relationships and interactions that impact each other and need to be accounted for.

With a multidimensional iterative model, it can be hard to intuitively determine, guess, or understand the downtime drivers. We know they are a subsystem or item that causes a technical system to be nonoperational, but truly understanding which systems and items those are, and if they change based on supply chain or utilization changes, proves tedious. SIMLOX allows the downtime driver results to be shown in both the results and reports. With the software update of 2024.1, any SIMLOX model in this and future versions will have the ability to showcase the downtime driver results.

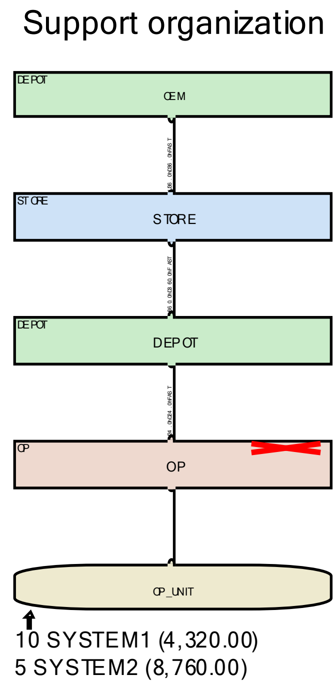
While the concept of downtime drivers is not new, the speed at which stakeholders can make decisions surrounding downtime drivers and modify optimization and simulation models to investigate the availability impacts from subsystems and items improves the collaborative and iterative decision-making process.

Model Scenario

The model for this use case is a pared-down radar system model. It contains two variants, 786 items, two depots, one store, and an operational location. There are a total of 15 systems deployed at the operational location.

- **Items:** Of the 786 items, 667 of them are universal to the radar (i.e., not variant specific). Every item is replaceable and repairable. There are 142 items that have a requirement of forced maintenance before the next mission. In addition, these items have a 90% probability for system effectiveness—meaning, 10% of the time, a failure on those items does not impact system availability or readiness—and a 90% probability of functional loss given a failure—meaning, 10% of the time there is not a functional loss when that item has a failure. There are 152 items that do not cause a functional loss and do not impact the system effectiveness. Prices for items range from $1 to $800,000, with the average being $11,727 and the mode and median being $1. Failure rates for these items have a range of 0.0008 to 1,000 failures per million operating hours. The mean is about 3 failures per million operating hours, the median is 0.3 failures per million operating hours, and the mode is 1 failure per million operating hours.
- **Maintenance:** Maintenance occurs at three levels. Level 1 is the removal and replacement of items from the system at the depot level. Level 2 is the complex removal and replacement of items at the depot level, as well as simple repairs of items at the depot level. The Level 3 is complex repairs at the OEM. More complex maintenance tasks take longer to complete.

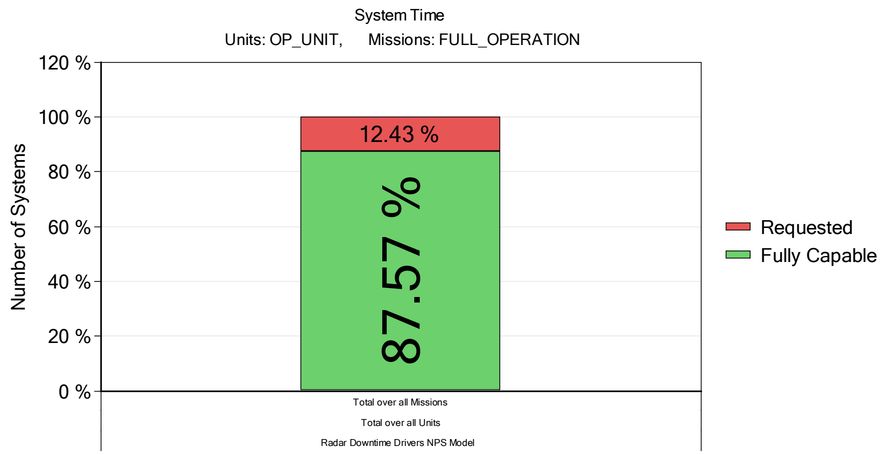
- **Support Structure:** This scenario has a basic support structure seen across many systems. There are two depots, which allow for storage and maintenance. The OEM is highest level, and it performs complex maintenance. In between the two depots, there is a store, which only allows storage of items. Items are reordered to both the OEM and the store. Below the store is the second depot, which is where intermediate maintenance occurs. Lastly, there is an operating station. At this station, maintenance and storage are not allowed. There is an operating unit assigned to this station and deployment of the systems assigned to it. Time to and from the operating station and the depot is 24 hours. Time to and from the depot and the store is 360 hours. Time to and from the store and the OEM is 36 hours.



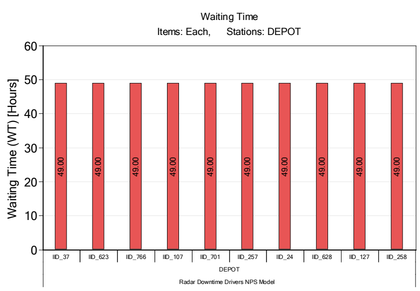
- **Missions and Deployment:** System 1, which is the variant 1, has 10 systems at the operational base. System 2 has five. Both systems are subject to a fixed 24-hour mission that requires one system. There are 64 of these missions within a year. That profile is repeated for each of the 3 years. The missions start at a randomized time and are distributed throughout the year randomly. Since the simulation is replicated 500 times, each replication has a different randomization; therefore, the overall results, which average utilization at a particular time, have factored in 500 possible randomized missions. Over the simulation period there are 192 missions that accumulate 4,608 system and mission hours.

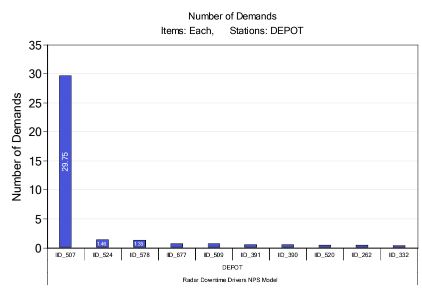
Model Results

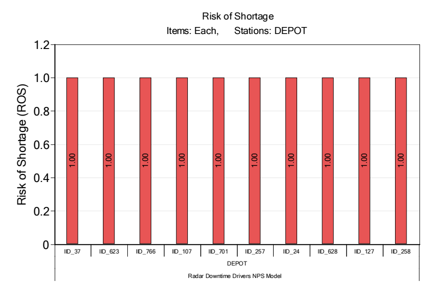
Overall, the simulation results show mission results—or how well the mission profile was carried out—to be 87.53% fully capable and 12.47% requested.



Historically, with SIMLOX results, we have looked at the item-specific information to try and identify downtime drivers. Graphs that include the number of demands, risk of shortage, and waiting time are a few that showcase what items are contributing to downtime of the system.







It is hard to tell from the item results graph exactly which items are downtime drivers and how much downtime they are causing.

Downtime Driver Results

The downtime driver results are subsystems or items-specific and are either time per a time period or accumulated time.

Table 1. System Downtime Caused by an Individual Item During an Interval of Time

	UNID	SID	OPMID	IID	STINT	ETINT	SYCDT
	Unit identifier	System identifier	Operational mode identifier	Item identifier	Start Time interval [Hours]	End Time interval [Hours]	System capability downtime [Hours]
1	OP_UNIT	SYSTEM1	<DEFAULT>	IID_507	0.00	24.00	0.0059
2	OP_UNIT	SYSTEM1	<DEFAULT>	IID_524	0.00	24.00	0.0013
3	OP_UNIT	SYSTEM1	<DEFAULT>	IID_611	0.00	24.00	0.00059
4	OP_UNIT	SYSTEM1	<DEFAULT>	IID_727	0.00	24.00	0.00089
5	OP_UNIT	SYSTEM1	<DEFAULT>	IID_305	24.00	48.00	0.00091
6	OP_UNIT	SYSTEM1	<DEFAULT>	IID_327	24.00	48.00	0.00038
7	OP_UNIT	SYSTEM1	<DEFAULT>	IID_484	24.00	48.00	0.0012
8	OP_UNIT	SYSTEM1	<DEFAULT>	IID_507	24.00	48.00	0.028
9	OP_UNIT	SYSTEM1	<DEFAULT>	IID_509	24.00	48.00	0.00049
10	OP_UNIT	SYSTEM1	<DEFAULT>	IID_524	24.00	48.00	0.0020
11	OP_UNIT	SYSTEM1	<DEFAULT>	IID_542	24.00	48.00	0.00038
12	OP_UNIT	SYSTEM1	<DEFAULT>	IID_553	24.00	48.00	0.0018

Table 2. Accumulated System Downtime Caused by an Individual Item During an Interval of Time

	UNID	SID	OPMID	IID	STINT	ETINT	SYCDT
	Unit identifier	System identifier	Operational mode identifier	Item identifier	Start Time interval [Hours]	End Time interval [Hours]	System capability downtime [Hours]
1	OP_UNIT	SYSTEM1	<DEFAULT>	IID_507	0.00	24.00	0.14
2	OP_UNIT	SYSTEM1	<DEFAULT>	IID_524	0.00	24.00	0.031
3	OP_UNIT	SYSTEM1	<DEFAULT>	IID_611	0.00	24.00	0.014
4	OP_UNIT	SYSTEM1	<DEFAULT>	IID_727	0.00	24.00	0.021
5	OP_UNIT	SYSTEM1	<DEFAULT>	IID_305	24.00	48.00	0.022
6	OP_UNIT	SYSTEM1	<DEFAULT>	IID_327	24.00	48.00	0.0091
7	OP_UNIT	SYSTEM1	<DEFAULT>	IID_484	24.00	48.00	0.030
8	OP_UNIT	SYSTEM1	<DEFAULT>	IID_507	24.00	48.00	0.82
9	OP_UNIT	SYSTEM1	<DEFAULT>	IID_509	24.00	48.00	0.012
10	OP_UNIT	SYSTEM1	<DEFAULT>	IID_524	24.00	48.00	0.079
11	OP_UNIT	SYSTEM1	<DEFAULT>	IID_542	24.00	48.00	0.0092
12	OP_UNIT	SYSTEM1	<DEFAULT>	IID_553	24.00	48.00	0.044

Note. The SYCDT adds up all time intervals prior and includes the current time interval.

Table 3. Accumulated System Downtime Caused by a Subsystem

	UNID	SID	OPMID	BDEID	STINT	ETINT	SYCDT	SYCDTA
	Unit identifier	System identifier	Operational mode identifier	Subsystem breakdown element identifier	Start Time interval [Hours]	End Time interval [Hours]	System capability downtime [Hours]	System capability downtime breakdown accumulated [Hours]
1	OP_UNIT	SYSTEM1	<DEFAULT>	RADAR_VARIANT_1	0.00	24.00	0.0087	0.0087
2	OP_UNIT	SYSTEM2	<DEFAULT>	RADAR_VARIANT_2	0.00	24.00	0.00	0.00
3	OP_UNIT	SYSTEM1	<DEFAULT>	RADAR_VARIANT_1	24.00	48.00	0.042	0.042
4	OP_UNIT	SYSTEM2	<DEFAULT>	RADAR_VARIANT_2	24.00	48.00	0.00	0.00
5	OP_UNIT	SYSTEM1	<DEFAULT>	RADAR_VARIANT_1	48.00	72.00	0.076	0.076
6	OP_UNIT	SYSTEM2	<DEFAULT>	RADAR_VARIANT_2	48.00	72.00	0.00	0.00
7	OP_UNIT	SYSTEM1	<DEFAULT>	RADAR_VARIANT_1	72.00	96.00	0.10	0.10
8	OP_UNIT	SYSTEM2	<DEFAULT>	RADAR_VARIANT_2	72.00	96.00	0.00	0.00

Note. In this case, the only subsystem in the model are the variants.

Since the only subsystems in this model are the variants, the focus for this use case is the items. Utilizing the information from these tables, you can identify what items are causing the most system downtime.

Modeling Impact

Showcasing the direct link between items and system downtime connects the dots between item decisions and their impacts. Assuming the input data for the simulation is accurate, and the maintenance tasks and support organization cannot be changed or modified, then you would be done with the downtime driver identification here. Deciding on what to do with this information will fall to decision-makers deciding if they want to invest in improving those items' failure rates or updating the support organization to improve the repair times.

If there is any uncertainty with the input data, or if you are modeling a system still in the design phase, then once the downtime drivers are identified, there needs to be a verification and inquiry as to why those might be the downtime drivers. In the example above, Item 507 causes an accumulation of over 1,400 hours of downtime. While this looks like a problem, upon further investigation, this item has a quantity of 500 in the systems, so the accumulated downtime for that item makes sense in this case. This item would then be counted as an outlier and not included in further decision-making addressing downtime drivers.

When there is room to adjust the support organization, maintenance capabilities, resource usage, or redesign items, then understanding and knowing the downtime drivers allows one to address those issues and find solutions. It is also important to see how varied usage of the support organization or resources, or maintenance capabilities can impact the downtime drivers. Identifying the overarching drivers amongst the model excursions leads one to believe that to improve the downtime for those items, an improvement on the item itself needs to be redesigned.

Quickly understanding how excursions or changes in the models impact or change the downtime drivers allows for better decision-making when supply chain issues arise or global events impact system readiness.

Data are crucial to the validity of a model or simulation; it is just as important that the results be interpreted and understood well so that decisions can be made quickly and effectively. Having explicit downtime driver results from SIMLOX reduces the risk of a poor decision being made by misinterpreted results. In addition, the ability to quickly visualize these results allows decision-makers to test possible solutions to overcome downtime and find the best option(s).

References

Systecon Group. (2018). *SIMLOX—Ensuring performance*. https://www.systecongroup.com/us/software/simlox-ensuring-performance

Systecon Group. (2020). *OPUS10—More than spare parts optimization*. https://www.systecongroup.com/us/software/opus10-more-spare-parts-optimization

Systecon North America. (2019, October 11). *US Navy common readiness model powered by Opus Suite* [Video]. YouTube. https://www.youtube.com/watch?v=hKDIJ9J_Rec

The Secret Sauce of Program Management Is the Best Defense to Mitigate Contract Risk

(Contract Management, Earned Value Management, and Agile Methodologies)

Christina Joseph—has an MS in Contract Management from the Naval Postgraduate School (NPS), and she was accepted into the NPS Acquisition Research Program (ARP) where her thesis was published in 2024. Christina has been dedicated to the Department of Defense (DoD) since June 2008. She began her career as a Branch Manager at the Defense Contract Audit Agency and is currently serving as a Supervisory Contract Specialist at the Defense Contract Management Agency (DCMA). Christina graduated with a bachelor's degree in accounting in 2008 and obtained her Certified Public Accountant (CPA) certification in 2010. Her professional journey is marked by a commitment to excellence in financial oversight and contract management within the DoD. Christina continues to advance her career goals with DCMA by leveraging her extensive experience and education to contribute to the agency's mission. [christina.joseph2.civ@mail.mil - 917.392.9116]

Symantha "Sam" Loflin—has an MS in Program Management and a certification in Advanced Acquisition studies from NPS, where she was a contributor to the 18th, 19th, 20th, and 21st Annual NPS Acquisition Research Symposium. She also holds a BS in Finance from the University of Houston. Sam is a Case Manager and Researcher at Tanner and Associates. She has more than 20 years of acquisition experience supporting DCMA, the DoD, NASA, and the military services. In 2020–2021, she served the United States as an acquisition program manager on the Coronavirus Task Force that focused on building the industrial base for personal protective equipment. Sam's career began at NASA, supporting the Space Shuttle, ISS, and the Constellation Programs in Houston. [symanthaloflin@gmail.com - 713.854.2933]

Abstract

The author has written this paper to defend and strengthen the use of government initiatives, industry, and academia risk mitigation measures that prevent divergence from successful Program Management (PM) with the framework of Contract Management (CM), and the integration of Earned Value Management (EVM), and Agile methodologies and practices. On December 14, 2016, Public Law No: 114-264, the "Program Management Improvement Accountability Act [PMIAA]" was signed into law (H.R. 114-637S.1550, 2016). This law was enacted to improve program and project management practices within the federal government by requiring agencies to conduct [document] annual portfolio reviews of "high risk" programs that the Government Accountability Office (GAO) identified. Additionally, the PMIAA establishing a Program Management Improvement Officer (PMIO), who will "assess the quality and effectiveness of program management" (2016). These measures will highlight the possibilities of future performance growth, increased demand, and technological advancements in the Defense Industrial Base (DIB) (DoD, 2022). Additionally, improvement in workforce acquisition career paths and skill levels (H.R. 114-637S.1550, 2016). At the onset, the "delivery of performance [will be] at the speed of relevance" (Mattis, 2018, p. 10).

Effective and efficient PM requires a solid foundation of knowledge and the framework of CM, with the integration of EVM, and Agile methodologies and practices. These disciplines will provide the capabilities required to maximize innovation, mitigate contract risk, and develop the workforce that supports the proper stewardship of taxpayer dollars.

In February 2022, the National Defense Industrial Association reported that U.S. national security interests are at risk given the declining health of the DoD's supply chain, surge readiness, and production capacity. It is essential to communicate and collaborate with all stakeholders to develop and grow the DIB and engage the workforce with the right people, processes, and tools at the right time.

Research Issue

Why is it imperative for acquisition professionals, policy-makers, and/or end users to have a thorough background and education of program management with the framework of CM, integration of EVM, and Agile methodologies and practices as a contract risk mitigation measure?

Research Results Statement

The results are clear that successful program management starts with the framework of CM, integration of EVM, and Agile methodologies and practices. It is your best defense to mitigate contract risk and to protect and grow the United States and the DIB. It takes a whole-of-government, industry, and acquisition professionals' approach that emphasizes collaboration, and the documentation of lessons learned to achieve and improve economic growth and protect national security.

The implementation and use of program management through the roles of the contracting officer, management, and functional specialist are essential to a fundamental assessment of the contractor's performance. In addition, combining the alignment of the framework of CM with the integration of data-driven performance tracking of EVM and Agile methodologies and practices will provide a more effective and efficient delivery of program goals. These disciplines enhance contract risk mitigation by providing information that can assist in controlling cost, increase schedule visibility, ensure contract requirements are met, and strengthen technical readiness.

The conformality of legal and regulatory procedures and the ability to engage the private sector to identify the best practices is essential to strengthen the use of program management. Investing in our workforce and our industry partners effectively and efficiently increases economic growth in the production of good, services, and materials in the United States that meet federal procurement needs.

Each year, the federal government increases the funding of developmental contracts as a measure to "Protect Sea, Air, and Space" (National Security Strategy, 2022 October 12). These measures aim to protect U.S. interests in developing emerging technologies, creating economic opportunities, and enabling climate surveillance, and to responsibly oversee and protect the sea, air, and space environment.

We must strengthen the acquisition professional's knowledge and effective execution of program management safeguards these measures. Having knowledgeable staff that know how and when to act will aid in protecting the U.S. interests in developing technologies will create economic opportunities to responsibly oversee the sea, air, and space environment. In addition, value is added to the taxpayer, federal government, and national security strategy.

Three Recommendations for Consideration

1. Increase training and educational opportunities for acquisition professionals on the effective and efficient measures of program management. Research shows that subject matter knowledge increases the likelihood of successful problem resolutions: "Experiential learning offers a way to ensure we are imparting not just rote learning and certifications but providing our people the knowledge, skills, and experience to effectively control the efforts we charge them to lead" (Pickar, 2020).

2. Establish collaboration policy and oversight tools that support program data gathering (H.R. 114-637S.1550, 2016).

3. Create a database to collect and disseminate evidence of lessons learned and best practices (GAO, 2019).

Introduction

This paper analyzes the whole-of-government approach to demonstrate that the program management secret sauce to mitigate contract risk is comprised of the framework of Contract Management (CM), with the integration of Earned Value Management (EVM), and Agile methodologies and practices.

The purpose of the investigation is to disclose what acquisition professionals, policy-makers, and/or end users need to know about the processes that are most appropriate to develop, acquire, and deliver these capabilities with emerging technologies within cost, schedule, and technical readiness to the warfighter.

Each year, the federal government increases the funding of developmental contracts as a measure to "Protect Sea, Air, and Space" (Biden-Harris Administration, 2022). These efforts aim to protect U.S. interests, while working with allies, academia, and our industrial base partners in advancing and developing new technologies that create economic opportunities and our shared security and prosperity.

An analysis of the research will reveal the key results related to the current and future measures that the government, military services, industry, and academia have endorsed PM with the framework of CM, and the integration EVM, and Agile methodologies and practices. Program Management measures are the best defense to analytically control cost, schedule, and technical implementation. When applied, effectively and efficiently PM will deliver timely capabilities and emerging technologies to the warfighter.

Methodology

The methodology will include a comparative analysis using literary research current and from the author's past Naval Postgraduate School's Acquisition Research Program Symposium research papers. Also, a variety of quantitative and qualitative information gathered from knowledgeable acquisition professionals related to Program Management (PM), framework of Contract Management (CM), Earned Value Management (EVM), and Agile methodologies and practices. This information will determine and reveal the best practices and contract risk mitigation measures related to program management that will build a stronger Defense Industrial Base.

Federal Risk Mitigation Measures - "Protect Sea, Air, and Space"

Buy American Act of 1933 and Buy American Act in 2021–2022

On March 3, 1933, during the Great Depression, Congress passed the Buy American Act (BAA),[1] and President Hoover signed it into law on his last day in office. When the BAA was enacted, it attempted "to protect domestic businesses and labor by establishing a price preference for domestic end products and construction materials in government acquisitions (Manuel, 2016, p. 1). In addition, the congressional oversight, by statute, requires agencies to submit a congressional report on procurement and compliance with the BAA that includes exceptions or trade agreement waivers. When solicitations contain the following clauses, federal government contracting officers who procure supplies are required to insert a FAR 52.225-2, Buy American Certificate and/or a FAR 52.225-6, Trade Agreements Certificate (TAA).

At the request of Senator Murphy, in December 2018, the GAO published their report on their review of four federal agencies' implementation of the Buy American Act. The GAO

[1] *The Buy American Act of 1933* (BAA) (41 U.S.C. ss8301-8305) requires federal agencies to purchase "domestic end products" and use "domestic construction materials" on covered above certain monetary thresholds (typically $10,000) performed in the United States. https://www.congress.gov/crs-product/R46748

reviewed 38 contracts from the Departments of Defense (DoD), Health and Human Services (HHS), Homeland Security (DHS), and Veterans Affairs (VA) and found that six contracts "inaccurately recorded waiver or exception information" (Woods, 2018). The GAO found that steps should be taken by the Office of Management and Budget (OMB) to improve Buy American Act data and by the agencies to improve implementation guidance and training on the Act (Woods, 2018, p. 1).

Given the government's past BAA reporting compliance errors, the Biden-Harris administration was dedicated to improving the BAA through policies and laws that include the Federal Acquisition Regulation(s) (FAR), Executive Order(s) (E.O.), and the establishment of the first Made in America Office (MIAO). The following highlights are a few of the improvements related to the BAA:

On January 25, 2021, Executive Order 14005, Section 4 (a) and Section 7 are noteworthy measures, Section 4 (a): the Director of the Office of Management and Budget shall establish the Made in America Office within the OMB (White House, 2021a). In April 2021, the Made in America Office (MIAO) opened to ensure "the future is made in America," strengthen domestic sourcing, and reduce the need for waivers.

Federal Contracting 1795

The Library of Congress (2025) provides an excellent source site, Federal Government Contracting: A Resource Guide. Most notable is the establishment of the Office of Purveyor of Public Supplies (1795) as it included Purveyor of Public Supplies in the Treasury Department and superintendent of Military Stores in the War Department (distributed supplies); the Commissary General of Purchase in the War Department (1812); and the Quartermaster's Department in the War Department. Over hundreds of years of contracting, the U.S. government has continued to enhance the framework of contract management (CM) over the years through statutory and regulatory measures.

Contract Management - Contractor Business Systems

Accounting System Administration as a Foundation for Risk Mitigation

Effective program management within the DoD requires reliable business systems that ensure fiscal integrity, compliance, and performance accountability—elements central to the "secret sauce" of delivering capabilities to the warfighter. Among these, the accounting system plays a foundational role in mitigating contract risk. As outlined in the Defense Federal Acquisition Regulation Supplement (DFARS) 252.242-7006, Accounting System Administration, contractors must meet 18 system criteria to ensure accurate financial data, appropriate cost segregation, and adherence to regulations like the Cost Accounting Standards (CAS) and Generally Accepted Accounting Principles (GAAP) (Defense Acquisition Regulations System [DARS], 2012). These criteria enable the DoD to monitor costs, validate performance, and safeguard government funds.

The accounting system serves as the financial infrastructure for collecting, managing, and reporting cost data, critical for both regulatory compliance and timely, data-driven decisions by program managers and contracting officials. The DoD's Fiscal Year 2025 budget request, released March 11, 2024, allocates $849.8 billion to strengthen military capabilities, with increased investments in research, development, test, and evaluation (RDT&E), underscoring the need for robust financial oversight in complex acquisition programs (DoD, 2024). Without such oversight, a deficient accounting system risks disapproval, payment withholds, or inaccurate cost reporting, undermining operational effectiveness and audit reliability (American Institute of Certified Public Accountants [AICPA], 2023; Joseph, 2024).

Joseph's (2024) analysis of 465 Defense Contract Audit Agency (DCAA) audit reports from the Defense Contract Management Agency (DCMA) Eastern Region (2020–2022) identified 47 disapproved accounting systems due to significant deficiencies (p. 38). Common issues included noncompliance with Federal Acquisition Regulation (FAR) Part 31, weak internal controls, and billing reconciliation problems, with frequent deficiencies in cost accounting information (criterion 15), unallowable costs (criterion 12), and segregation of direct and indirect costs (criterion 2) (Joseph, 2024, p. 53). These findings highlight accounting systems' pivotal role in supporting or hindering risk mitigation in program execution.

Alignment with Internal Control Standards and Policy Reform

To enhance clarity and enforceability, the DFARS 252.242-7005 clause was revised as of January 17, 2025, replacing "significant deficiency" with "material weakness," aligning DoD standards with the AICPA's Generally Accepted Auditing Standards (GAAS) (AICPA, 2023; DARS, 2025). This shift, enacted via Section 806 of the National Defense Authorization Act (NDAA) for FY 2021, adopts a GAAS-consistent definition of "material weakness" as a deficiency posing a reasonable possibility of material misstatement, fostering consistency between government and industry practices (NDAA, 2020).

Separately, the Committee of Sponsoring Organizations of the Treadway Commission's (COSO) Internal Control—Integrated Framework (2013) informs efforts to streamline oversight, such as the Section 809 Panel's Recommendation 72 to reduce the 18 DFARS criteria to seven (Committee of Sponsoring Organizations of the Treadway Commission [COSO], 2013; Section 809 Panel, 2019). These reforms enhance contractor business systems' auditability and transparency. The Section 809 Panel's (2019) proposal aligns with COSO principles, aiming to maintain oversight while easing compliance burdens, particularly for small businesses (Section 809 Panel, 2019). Joseph (2024) supports this, noting that four of five disapprovals involved small businesses struggling with DFARS complexity (p. 81). Such reforms could improve compliance and reduce risk across the Defense Industrial Base (DIB).

Integration with Program Management Disciplines

The accounting system integrates with program management frameworks like Contract Management (CM), Earned Value Management (EVM), and Agile acquisition practices. In CM, it supports compliance verification, invoice validation, and audit preparation. For EVM, reliable cost data drives performance metrics like the Cost Performance Index (CPI) and Schedule Performance Index (SPI), enabling real-time program health assessments (Defense Acquisition University [DAU], 2023). Agile acquisition, emphasizing flexibility and iterative delivery, relies on accurate, timely financial reporting to adjust resource allocation as priorities evolve.

This integration shines in EVM, where DFARS criteria like labor distribution (criterion 10), general ledger control (criterion 5), and interim cost determination (criterion 11) ensure cost data accuracy for CPI and SPI calculations (DARS, 2012; DAU, 2023). Joseph (2024) found contractors with stronger accounting systems were more likely to resolve deficiencies via Corrective Action Plans (CAPs), achieving a 69% DCMA-DCAA agreement rate on corrective action adequacy (p. 80). This positions the accounting systems as active tools for managing cost, schedule, and performance risks, beyond mere compliance.

Agile Methodologies and EVM Integration

On December 13, 2023, the GAO issued *Agile Assessment Guide*: *Best Practices for Adoption and Implementation*. Agile is a description of iterative, incremental software development methods. Agile concepts and methodologies include frameworks and practices known as Lean, Kanban, Scrum, Feature Driven Development (FDD), Extreme Programming (XP), DevOps, and others. These methodologies are central to the Agile principles of adaptability, flexibility, collaboration to deliver stakeholder value, continuous improvement.

The Agile Assessment Guide provides a clear and concise best practices approach for program monitoring and control when incorporating and applying earned value management on Agile programs (GAO, 2023). The Section 809 Panel, of the National Defense Authorization Act for Fiscal Year 2016 (Pub. L. No. 114-92), recommended eliminating EVM requirements for Agile programs (Section 809 Panel, 2018).Additionally, the Panel recommended that Program Executive Officers (PEOs) be authorized to approve "appropriate project monitoring and control methods" (Wahidi, 2022). The bottom line is that, for EVM to effectively and efficiently integrate with Agile, the program office must tailor EVM into the total program management methodology.

Conclusion

Accounting system administration anchors DoD program management, ensuring cost control, compliance, and audit readiness. As acquisition evolves with emerging technologies, robust accounting systems are vital for mitigating contract risk, enabling the warfighter to receive cutting-edge capabilities on time and within budget. Strengthening these systems through policy reform, training, and the framework of CM, the integration of EVM, and Agile methodologies and practices enhances contractor performance and government oversight. These management practices will highlight the possibilities of future performance growth, increased demand, and technological advancements in the Defense Industrial Base (DIB) and in professional acquisition skill levels. At the onset, the "*delivery of performance [will be] at the speed of relevance*" (Mattis, 2018, p. 10).

List of References

Abba, W. F. (2017). The evolution of earned value management. *Defense AT&L Magazine.*

American Institute of Certified Public Accountants. (2023). *Statement on auditing standards No. 145: Understanding the entity and its environment and assessing the risks of material misstatement*. https://www.aicpa.org/resources/download/statement-on-auditing-standards-no-145

Assistant Secretary of the Navy for Research, Development, and Acquisition. (2022, April 8). *Department of the Navy implementation of the defense acquisition system and the adaptive acquisition framework* (SECNAVINST 5000.2G). Department of the Navy.

Biden-Harris Administration. (2022, October 12). *National security strategy*.

Committee of Sponsoring Organizations of the Treadway Commission. (2013). *Internal control—Integrated framework: Framework and appendices*. https://egrove.olemiss.edu/cgi/viewcontent.cgi?article=1769&context=aicpa_assoc

Defense Acquisition Regulations System. (2012). DFARS 252.242-7006, Accounting system administration. https://www.acq.osd.mil/dpap/dars/dfars/html/current/252242.htm#252.242-7006

Defense Acquisition Regulations System. (2025). DFARS 252.242-7005, Contractor business systems. https://www.acq.osd.mil/dpap/dars/dfars/html/current/252242.htm#252.242-7005

Defense Acquisition University. (2023). *Defense acquisition guidebook: Chapter 1—Program management*. https://www.dau.edu/sites/default/files/2023-09/DAG-CH-1-Program-Management.pdf

Department of Defense. (2022, June 8). *Operation of the adaptive acquisition framework* (Department of Defense Instruction 5000.02). https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500002p.PDF

Department of Defense. (2024, March 11). *Department of Defense releases the president's fiscal year 2025 defense budget* [News release]. https://www.defense.gov/News/Releases/Release/Article/3703410/department-of-defense-releases-the-presidents-fiscal-year-2025-defense-budget/

DFARS 252.242-7006, Accounting System Administration (2012, February). https://farclause.com/FARregulation/Clause/DFARS252.242-7006_Basic-accounting-system-administration#gsc.tab=0

GAO. (2019, December*). Report to congressional requestors. Evidence-based policymaking. Selected agencies coordinate activities, but could enhance collaboration.* https://www.gao.gov/products/gao-20-119

GAO. (2023, December 13). *Agile assessment guide: Best practices for adoption and implementation* (GAO-24-105506). https://www.gao.gov/assets/d24105506.pdf

H.R. 114-637S.1550 - Program Management Improvement Accountability Act (2016, December 14). https://www.congress.gov/bill/114th-congress/senate-bill/1550

Joseph, C. (2024). *Analysis of DFARS 252.242-7006 accounting system administration significant deficiencies* [Master's thesis, Naval Postgraduate School]. Acquisition Research Program Sponsored Report Series. https://dair.nps.edu/handle/123456789/5234

Library of Congress. (2025, April 18). *Research guides, federal government contracting: A resource guide*. https://guides.loc.gov/federal-government-contracting/understand-past

Mattis, J. (2018). *Summary of the 2018 national defense strategy of the United States of America*. Department of Defense. https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf

National Defense Authorization Act for Fiscal Year 2021, Pub. L. No. 116-283, § 806, 134 Stat. 3388 (2020). https://www.congress.gov/bill/116th-congress/house-bill/6395/text

National Defense Industrial Association. (2022, February). *Vital signs 2022: The health and readiness of the defense industrial base*. https://www.ndia.org/-/media/sites/ndia/policy/vital-signs/2022/vital-signs_2022_final.pdf

Pickar, C. (2020). *Learning from experience: Acquisition professional education for this century* (SYM-AM-20-070). Naval Postgraduate School, Acquisition Research Symposium.

Section 809 Panel. (2018, January). *Report of the advisory panel on streamlining and codifying acquisition regulations, Volume 1 of 3*.

Section 809 Panel. (2019). *Report of the advisory panel on streamlining and codifying acquisition regulations: Volume 3 of 3*. https://discover.dtic.mil/wp-content/uploads/809-Panel-2019/Volume3/Sec809Panel_Vol3-Report_Jan2019_part-1_0509.pdf

The Buy American Act of 1933 (BAA) (41 U.S.C. ss8301-8305). https://www.congress.gov/crs-product/R46748

Wahidi, R. S. (2022, December). *Charting progress in the software acquisition pathway* (NPS-AM-23-05). Naval Postgraduate School, Acquisition Research Program. https://dair.nps.edu/handle/123456789/4785

White House. (2021, January 25). *Ensuring the future is made in all of America by all of America's workers was signed and Section 4 (a): Updating and centralizing the Made in America waiver process* (Executive Order 14005).

Woods, W. T. (2018, December). *Buy American Act: Actions needed to improve exception and waiver reporting and selected agency guidance* (GAO-19-17). Government Accountability Office.

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net