Recommendation 89: Direct DoD to consolidate or eliminate competing data architectures within the defense acquisition and financial system.

Problem

The proliferation of different data architectures throughout the acquisition and financial system leads to countless marginal inefficiencies in DoD and the rest of government. Collectively, these inefficiencies lead to poor interoffice communication, entrenchment of institutional siloes, lower adaptability of labor, and damaged workforce morale. One senior acquisition leader characterized the large number of data architectures as a Tower of Babel preventing communication across the organization.[1]

Virtually all stakeholders in the defense acquisition system—requirements developers, program managers, congressional committees, contracting personnel, financial managers—use different financial data architectures, many of which are incompatible with one another. Data architecture proliferation is an extremely challenging problem, with both technological and political aspects. Too many preexisting stakeholders are wedded to their own ways of doing business and unwilling to change their approaches to accommodate other stakeholders.

Background

Recent decades have seen a proliferation of data systems in federal acquisition. These data systems contain information on requirements development, appropriations requests, budget allocations, program management, contracting, contract audits, disbursement of funds, and many other functional areas that are core parts of the acquisition process. With the proliferation of data systems, a proliferation of data architectures has also occurred.[2] Even within individual organizational jurisdictions or functional communities, multiple data architectures cause confusion in many cases.

These different architectures effectively constitute different languages, inhibiting both comparative data analysis and data tracking from one transaction to another. Because of these different languages, useful data analysis requires a large number of human and machine translators throughout the process. Many of the machine translators are aging systems to which few people pay attention until they break or produce obvious errors. Such errors lead to unnecessary dysfunction, added maintenance costs, and small data errors that can persist for years unnoticed by analysts. Many of the human translators serve as the sole providers of niche expertise in their workplaces. This situation leads to inefficiency in the form of work backlogs, long wait times, and potential labor market distortions.

Misallocation of Skilled Labor

Each year, DoD employees and support contractors spend an enormous amount of time performing the work required to interpret nonstandardized data across different functional communities. It is

---

[1] Retired Air Force general officer, discussions with Section 809 Panel, July 2018.

[2] A 2017 study of 21 acquisition and financial data systems found that one of the overarching problems was inconsistency of technical terminology and data formats across systems: "The same term can have different meanings in different acquisition systems, which makes analyses across systems particularly challenging." See Megan McKernan et al., *Issues with Access to Acquisition Data and Information in the Department of Defense: Doing Data Right in Weapon System Acquisition,* RAND Corporation, 2017, 41, accessed July 21, 2018, https://www.rand.org/pubs/research_reports/RR1534.html.

impossible to calculate an exact number of annual labor hours associated with this type of work, but it is likely in the millions.

Many of these people are among the most highly-skilled data science professionals working for DoD. If acquisition and financial activities were to adopt more uniform data architectures, it would free up a large volume of skilled labor for more substantive tasks, such as data analysis and strategic planning.

Data Analytics Versus Data Architecture

In recent years, many proposals in the Legislative and Executive Branches have focused on improving the quality of data analytics in DoD.[3] These are commendable initiatives, as the quality of governmentwide financial data analytics clearly needs improvement.[4]

Improving the quality of data analytics in DoD is mostly about personnel, not policy. Academic studies and senior-level meetings will not build these capabilities. The most important part of improving DoD data analytics is ensuring that creative and competent data science professionals choose to work for the organization. Policy changes may be able to affect workforce recruitment and retention somewhat, but workforce quality will improve substantially in the long run only if capable people are given the leeway to innovate and are committed to DoD's mission. Poor data analysis is a personnel issue; there are no miracle cures for such problems.

What is Data Architecture?

In its broadest sense, the term data architecture refers to the way information is organized. The term *data architecture* often refers to the organization of formally structured relational databases with data elements that are standardized throughout the database. A defense acquisition example would be the contract-writing systems run by the Military Services, which track and report data on contract and related actions using standardized formats.

The term data architecture could also be used to refer to unstructured datasets such as large repositories of documents written in free-form prose. A defense acquisition example would be the Acquisition Information Repository (AIR), a web-based system allowing access to a large number of program management documents.[5]

In the context of this recommendation, data architecture refers to the structure of information that DoD and other stakeholders use to categorize different parts of the acquisition system. These parts include, but are not limited to, the following.

---

[3] As examples, see Section 913 in the FY 2018 defense authorization, "Establishment of set of activities that use data analysis, measurement, and other evaluation-related methods to improve acquisition program outcomes." FY 2018 NDAA, Pub. L. No. 115-91 (2017). Also see U.S. Department of the Army, *Enterprise Data Analytics Strategy for Army Business 2018-2022,* accessed July 10, 2018, https://www.army.mil/e2/downloads/rv7/professional/enterprise_analytics_strategy.pdf.
[4] GAO, *Contracting Data Analysis: Assessment of Government-wide Trends, GAO-17-244SP,* March 2017, accessed July 10, 2018, https://www.gao.gov/assets/690/683273.pdf.
[5] AIR provides access to program management documents such as Acquisition Strategies, Test and Evaluation Master Plans, and Analyses of Alternatives. It is designed in part to help program managers write required acquisition documents by using documents from other successful programs as templates.

- The requirements development community categorizes capabilities based on the needs of current and future missions.
- The programming community compiles lists of programs that are needed to meet required capabilities.[6]
- The budgeting community compiles lists of line items that will require congressional appropriations to fund required programs.
- Congress enacts appropriations bills into law using a much more aggregated version of the budgeting community's data architecture. It also publishes clarifying reports using the same data architecture as the budgeting community.
- The financial management community releases funds to the program management community, using additional data architectures meant to translate between appropriations accounts and the program-level perspective.
- The program management community uses a variety of data architectures to manage funds based on appropriation availability.
- The contracting community uses several different data architectures to manage contract solicitations and awards. Organizations such as the General Services Administration (GSA) and Small Business Administration use their own data architectures for tracking certain aspects of DoD contracting.
- The accounting and contract close-out communities use their own data architectures to track contract finances. The Department of the Treasury is also involved in final disbursement of funds and has data architectures to track that process.

Each of these broadly defined data architectures contains a variety of data elements. For example, the top-level data architecture used for public reporting of contract awards has roughly 200 data elements in 14 subcategories. These elements include unique transaction identification numbers, dates of contract signing, dollar obligations, agencies, legislative mandates, and many other data elements.[7]

To some extent, mismatches between different architectures may be inevitable. When translating mission requirements into programs, for example, changing technologies and external dynamics may require that individual program subcategories be reorganized; however, minimizing the frequency and scope of these reorganizations is important to ensure that communication across different functional communities does not require an excessive amount of translator middlemen.

[6] The term *programming community* in this context refers to the resource managers and other personnel who specialize in translating technical warfighting requirements into specific acquisition programs.
[7] "FPDS-NG Data Dictionary" (Federal Procurement Data System—Next Generation documentation), GSA, accessed July 18, 2018, https://fpds.gov/wiki/index.php/V1.5_FPDS-NG_Data_Dictionary.

History

The data architectures of the modern acquisition system date back at least to the mid-1900s. In many cases they evolved in iterative, organic ways that make it difficult to pin their development to specific years.

Much of the modern budgeting data architecture can be traced to the National Security Act of 1947 and subsequent amendments. It empowered the Secretary of Defense to,

> *Supervise and coordinate the preparation of the budget estimates of the departments and agencies comprising the National Military Establishment; formulate and determine the budget estimates for submittal to the Bureau of the Budget; and supervise the budget programs of such departments and agencies under the applicable appropriation Act.*[8]

An amending law in 1949 established the office of the DoD Comptroller and directed the individual holding this position to "establish uniform terminologies, classifications, and procedures" in matters related to "budgetary and fiscal functions as may be required to carry out the powers conferred upon the Secretary of Defense."[9]

The data architecture that underpins the defense budget was highly unstructured for most of U.S. history. In appropriations laws from the early 1900s, congressional appropriations were enacted at regular intervals throughout the year as needed. Funding categories were nonstandardized and ad hoc, introduced based largely on congressional interest.[10] In many cases, individual laws were enacted to provide funding for specific projects, pensions, or people.[11]

After the 1949 legislation, however, budget estimates began to adopt a format roughly identical to that used today. The FY 1953 defense appropriations act laid out a now-familiar data architecture for appropriations accounts. To use the Army accounts as an example, accounts included

- Military Personnel, Army;
- Maintenance and Operations, Army;
- Procurement and Production, Army; and
- Research and Development, Army.[12]

---

[8] Section 202(a)(4) of the National Security Act of 1947, Pub. L. No. 80-253 (1947). The Bureau of the Budget was the predecessor to the modern-day Office of Management and Budget.

[9] Section 11 of the National Security Act Amendments of 1949, Pub. L. No. 81-216 (1949).

[10] For example, see An Act Making appropriations to supply urgent deficiencies in appropriations for the fiscal year nineteen hundred and ten, and for other purposes, Pub. L. No. 61-62 (1910). The law included three Navy appropriations: $300,000 for dredging an entrance channel at Pearl Harbor, HI; $48,136.46 for repair of the Mare Island, CA, shipwright's shop; and $10,000 for heating and lighting the men's band quarters at the U.S. Naval Academy.

[11] Virtually all legislation enacted today takes the form of public law, but Congress used to enact dozens or even hundreds of private laws per year. These frequently took the form of "An Act For the relief of" or "An Act Granting a pension to" individuals. As an example, a U.S. government mail carrier allegedly stole $208.44 out of a letter intended for Mr. Jacob Pickens of Neosho, Missouri. Congress enacted a law in 1906 reimbursing Mr. Pickens via funding from the Treasury. See Private Law 59-1526 (Chap. 1628, "An Act For the relief of Jacob Pickens"), enacted April 14, 1906.

[12] From Department of Defense Appropriations Act, 1953, Pub. L. No. 82-488 (1952). For comparison with earlier appropriations law data architecture, see Fourth Supplemental Appropriation Act, 1951, Pub. L. No. 81-43 (1951).

Within each of these accounts are more detailed budget line items that lay out in detail the intended expenditures for each account.

The data architecture behind contract awards dates back at least to the mid-1970s, when the Federal Procurement Data System Committee laid out a standardized set of data elements applicable to DoD and major civilian agencies.[13] The data architecture the committee built was used as the basis for the Federal Procurement Data System (FPDS), evolved over the course of subsequent decades, and continues to be used and modified in FPDS's successor IT system as of 2018.

The modern data architecture behind DoD's contract accounting systems can be traced back at least as far as the 1950s, when the Mechanization of Contract Administration Services system was developed. The system was still in use as of 2018, supporting the business processes for about $1.6 trillion in contract obligations and entitlements.[14] These are some of the most prominent and well-known data architectures that form the backbone of acquisition and financial data collection for executive agencies. Many other data architectures have gradually developed in response to various organizational needs.[15]

These data architectures largely evolved organically over time, independent of one another. They have been modified over the course of multiple generations by executive agency officials, business process managers, and IT system developers. Laws and regulations are sometimes written with a policy objective in mind but without substantial consideration for the reorganization of data architectures required for implementation. Private-sector companies rarely build commercial off-the-shelf enterprise resource planning software to comply with U.S. executive agency rules, meaning that customized systems must frequently be built to convert commercial data architectures into federal government-compliant architectures. The end result is a system in which relatively few data architectures are purpose-built to be compatible with each other.

Below are several examples of specific data architectures that are used in the DoD acquisition system, as well as examples of incompatibilities.

Requirements

Joint Capability Areas (JCAs) are one of the predominant data architectures used by the requirements community in DoD. JCAs are laid out in the Joint Capabilities Integration and Development System manual, which provides examples of each category.[16] They are described as collections of attributes and include categories such as force support attributes, battlespace awareness attributes, and force

---

[13] GAO, *The Federal Procurement Data System—Making It Work Better,* April 18, 1980, accessed July 16, 2018, https://www.gao.gov/assets/130/129310.pdf. Also see Section 6(d)(5) of the Office of Federal Procurement Policy Act, Pub. L. No. 93-400 (1974).

[14] Discussion between Chairman Jason Chaffetz and DoD CIO Terry Halvorsen during testimony before House Committee on Oversight and Government Reform, *Federal Agencies' Reliance on Outdated and Unsupported Information Technology: A Ticking Time Bomb,* May 25, 2016, 83, accessed July 16, 2018, https://www.gpo.gov/fdsys/pkg/CHRG-114hhrg23644/pdf/CHRG-114hhrg23644.pdf.

[15] See, for instance, the Product and Service Code (PSC) system, which is used for policy purposes to determine whether an acquisition is a *product* or *service*. The PSC data architecture is jointly built by DoD, GSA, and the NSF. Also see the North American Industry Classification System (NAICS) codes and SBA Small Business Size Standards, which are used to determine whether a potential vendor is or is not a small business. NAICS codes are built by NAFTA member country agencies.

[16] DoD, *Manual for the Operation of the Joint Capabilities Integration and Development System (JCIDS),* Appendix A: JCA Specific Examples, modified December 18, 2015, accessed November 8, 2018, http://www.acqnotes.com/wp-content/uploads/2014/09/Manual-for-the-Operationsof-the-Joint-Capabilities-Integration-and-Development-System-JCIDS-18-Dec-2015.pdf.

application attributes.[17] These categories of attributes do not match the data architectures used among the offices that develop budget requests to send to Congress. Indeed, the attributes used by the requirements development community do not match any data architectures used by any community for the vast majority of the acquisition process. Arguably, the only matching data architectures are those used to categorize solutions available to warfighters when carrying out missions.

Budget Requests

When DoD and other agencies submit budgets to Congress each year, they must convert the data architecture of requirements into the data architecture of the President's Budget Request (PBR). The PBR's data architecture is largely mapped to the architectures of congressional regular appropriations bills.[18] PBR documents for DoD contain a top-level data architecture with individual line items representing appropriations accounts. Within these accounts are data elements called budget activities. Within the budget activity data elements are further subelements called budget subactivities, with additional layers of lower-level data elements.[19] These data layers largely map to those presented in defense authorization bills and regular appropriations bill conference report joint explanatory statements. They do not directly map to most of the data architecture used by the requirements community to define what the military needs to accomplish its missions.

Appropriations

Appropriations data architecture consists of several accounts that appear in roughly the same format for each annual bill. These accounts contain dollar amounts that agencies are to spend, as well as constraints on how they are to spend them. The amounts defined in appropriations accounts legally bind agencies to spend the amounts indicated, with some exceptions for flexibility.[20] Within each of the top-level appropriations accounts are lower-level data elements that appear in appropriations bill conference report joint explanatory statements. These lower-level data elements describe how much agencies are expected to spend on each line item indicated. They are not legally binding to the same degree as appropriations accounts.

Budget Allocations and Allotments

Once Congress approves funding, the data architectures used by agency comptrollers closely resemble those used by the budget request and appropriations communities. Distribution of funding throughout DoD is based on appropriations accounts at the congressional level, apportionment at the OMB level, allocation at the DoD Comptroller level, suballocation at the Military Service comptroller level,

---

[17] "Joint Capability Area (JCA)," Defense Acquisition Glossary, accessed January 29, 2018, https://www.dau.mil/glossary/pages/2104.aspx.

[18] *Regular appropriation* is a term of art in fiscal law, referring to 12 specific bills per year that overlap with appropriations subcommittee jurisdiction. The defense appropriations bill is one of these 12. Regular appropriations bills are distinct from other types of congressional appropriations laws, such as continuing resolutions and supplemental appropriations. Unlike regular appropriations bills, these bills do not always have a uniform, standardized, recurring data architecture.

[19] For example, see FY 2019 DoD budget request for the Procurement appropriations accounts. Within the Aircraft Procurement, Army appropriations account are the budget activities Aircraft, Modification of Aircraft, Support Equipment and Facilities, and Undistributed. Within the Aircraft budget activity are budget subactivities Rotary and Fixed Wing. Within the Rotary budget subactivity are additional data element layers indicating type of aircraft and funding approach.

[20] For example, DoD reprogramming under the general transfer authority and DoD 7000.14-R Financial Management Regulation, Volume 3, Chapter 6.

allotment at the major commands level, and suballotment at the program management level.[21] These data architectures do not precisely map to those used by congressional appropriators. Data elements within a Military Service's budget structure originate in accounts from multiple regular appropriations bills.[22]

Program Management

The data structure of acquisition program management does not conform to those of budgeting and appropriations. A program manager reports to a program executive officer, who ultimately reports to a service acquisition executive and the Office of the Secretary of Defense. Virtually all of these tiers of command must commit funds from multiple appropriations accounts to do their job. There is substantial data architecture misalignment between the budgeting community and the program management community.

Contract Awards

Several unstructured data architectures are used to describe various phases of the contracting process prior to award. For example, the contract solicitations posted to Federal Business Opportunities (FBO) generally consist of a few standardized data elements and several text documents or spreadsheets describing solicitation requirements. The financial data architecture of government contracts is largely based on contract line-item numbers (CLINs). CLINs describe what is being purchased as well as the appropriations identity and fiscal year.[23] It is common for individual contracts to use CLINs from multiple appropriations accounts to meet multiple technical requirements. There is data architecture misalignment between the contracting and requirements communities, as well as between the contracting and budgeting communities.

When a contract award is made, there is a standardized data architecture in place for reporting the award. Agencies collect data related to the award, and some 200 of these standardized data elements are in most cases reported publicly.[24] This standardized data structure is largely built based on legal and regulatory requirements, and bears little resemblance to data architectures used by the requirements community, budgeting and appropriations community, or program management chains of command.

Data Architecture Mismatch Effects

When data architectures are not aligned across functional communities, those functional communities find it much more difficult to communicate with one another. This situation is partly because of the lack of a shared lexicon. It is also partly because more complicated data architectures necessitate involving more stakeholders in decision making. This leads to deep inefficiencies in the acquisition process. As an example, a military department might determine that instead of a hardware IT solution, a mission would be more effectively accomplished via a services contract solution accompanied by

---

[21] As they apply to defense acquisition, these processes are laid out in DoD 7000.14-R Financial Management Regulation, Volume 14, Chapter 1. For a simplified overview of these processes, see William Fast, "Budget Execution 101," presentation at Defense Acquisition University, June 3, 2010.
[22] Congress enacts about 12 regular appropriations bills each year. The majority of military department budgets come from the *Department of Defense* appropriations bill. Some also comes from the *Military Construction, Veterans Affairs, and Related Agencies* bill.
[23] See FAR Subpart 4.10 for detailed CLIN requirements.
[24] Contract award data publicly available via Federal Procurement Data System or USA Spending.

changes to internal business processes. If the hardware solution is already a program of record in the budget, this transition process will take years. The Military Service would have to reorganize its budget request, wait for Congress to appropriate funding, reorganize or close down the program office, and get the contracting community to obligate the funding on a new contract. If data architectures throughout the acquisition system were uniformly aligned to missions, the department would simply be able to reallocate the already-appropriated funding from the hardware solution to a service and business process solution. It would eliminate the need for much of the waiting that occurs throughout this process.

Mismatched data architectures also impede oversight by limiting transparency. When different communities speak different data languages, it can be difficult for even a willing executive agency to provide understandable information to an inspector general, auditor, or congressional committee. For example, data on contract transactions is collected and reported in a way that is out of sync with both the needs being met and the budgeting process. Determining the appropriations account that funded a given transaction is challenging at best.

It can also be extremely difficult, using standardized contract transaction data, to determine what problem a contract was intended to solve. This situation is largely because the data architecture of product, service, or industry codes does not provide meaningful detail.[25] This situation also limits agencies' ability to engage in strategic acquisition planning.

In addition to misalignment across functional communities, data architecture misalignment across jurisdictions can reduce efficiency. GAO noted in 2018 that "DFAS, DLA, and WHS differ in how they measure and report their performance data, which results in inconsistent information and limits customers' ability to make informed choices about selecting a human resources service provider to meet their needs."[26] These and many other architectural issues prevent oversight officials and strategic planners from understanding of the inner workings of acquisition and financial data.

Problems with Data System Consolidation

Consolidating different data systems may be part of a solution, because it can force people to adopt uniform data architectures across different datasets. In too many cases, however, data system consolidation appears to simply take the form of new, more aesthetically pleasing interfaces between existing systems. This approach generally leads to a more uniform top-level data architecture. The original data architectures, however, continue to exist and require maintenance and updates. Maintaining these architectures is made more difficult because the popularity of top-level architecture means there are fewer people who understand the underlying architectures well enough to manage the software systems in which the architectures live. In a best-case scenario, this situation may lock DoD into paying economic rents to system integrators for decades to come. In a worst-case scenario, it may

---

[25] The data architecture used for identifying contract product and service categories is the Product and Service Code (PSC) system. For some areas PSCs can be fairly detailed, such as PSC Q501, "medical services – anesthesiology." For other PSCs, the lack of detail makes them effectively meaningless for understanding why something is being purchased. For example, PSC 7030, "information technology – software," can encompass anything from logistics to human resources to health care.
[26] GAO, *Defense Management: DOD Needs to Address Inefficiencies and Implement Reform across Its Defense Agencies and DOD Field Activities, GAO-18-592*, September 2018, 19, accessed September 13, 2018, https://www.gao.gov/assets/700/694333.pdf.

increase cybersecurity risks due to the limited number of in-house personnel who are knowledgeable enough to address problems as they arise.

Other Issues: Data Architecture Ownership

Ownership is an important issue in acquisition and financial data management. In the past, robust debates have centered around whether data architectures should be government-owned or privately-owned. Common complaints about government-owned architectures involve costs associated with maintenance and documentation, as well as the need to keep systems updated to reflect the real world they are meant to represent.

Adopting privately owned data architectures and incorporating them into government systems may in some cases provide better-quality processes in the short term. This approach, however, can effectively lock the government into an arrangement under which it must pay in perpetuity for continued access to a monopoly translator service between two data systems. When sufficiently integrated into government processes, such a translation service can be difficult to replace with a government-controlled alternative.

Case Study:
Government Ownership of Company Identifier Data Architecture

Around 1960, DoD began using a data architecture called the H4/H8 system for tracking the corporate identities of contractors in North Atlantic Treaty Organization countries.[27] Essentially, the system allowed people to connect the data on a contract with data indicating a company's identity and ownership.

In 1975, a congressionally established committee identified several problems with the H4/H8 system. Among them was the need for the government to publish data manuals and maintain up-to-date records of corporate name changes, mergers, and acquisitions.[28]

The committee noted that a private company might be better able to provide high-quality, up-to-date information on these changes to corporate identity. It expressed concerns about "whether it is appropriate to have a sole source contract" and "what happens if the contractor goes out of business."[29] The committee and DoD stakeholders at the time also noted higher costs associated with a private sector solution. GAO analysts assessed that potential improvements to analytical capabilities did not justify transferring from a government-controlled to a privately-owned data architecture for tracking corporate identifiers.[30]

---

[27] Based on GAO statement that in 1975, DoD "had been using its own system for 16 years." GAO, *The Federal Procurement Data System—Making It Work Better,* April 18, 1980, 17, accessed November 8, 2018, https://www.gao.gov/assets/130/129310.pdf.
[28] GAO, *The Federal Procurement Data System—Making It Work Better,* April 18, 1980, 16, accessed November 8, 2018, https://www.gao.gov/assets/130/129310.pdf.
[29] Ibid.
[30] Ibid, iii.

Case Study:
Government Ownership of Company Identifier Data Architecture

Despite these concerns, the committee recommended in 1975 that the government adopt the proprietary data architecture.[31] In the late 1990s, the proprietary system was formally incorporated into the Federal Acquisition Regulation.

In 2012, GAO published another report on the issue of corporate identifier data architecture. The report noted that in the span of a decade, contract costs had increased roughly 1,800 percent.[32] GAO attributed this increase in part to an "effective monopoly" that "contributed to higher costs."[33] Switching to a government-owned system would resolve this issue, but for technical reasons, this was not an option in the near term. By the time of GAO's report, the privately-owned data architecture had "become an integral component in how government data systems operate."[34]

Other Issues: Black Box Data Elements

In recent years, a cottage industry has arisen wherein analytics companies sell their analyses of government data to the government. There is nothing inherently wrong with this model. In many cases, analytics firms may provide insights, recommendations, or dashboard visualizations that add value to government operations.

In at least some cases, however, these companies' products essentially constitute new data architectures in their own right. Several firms perform automated analyses of existing data elements in federal contract award and related data. Through proprietary computer code—an opaque black box—these companies create new, proprietary data elements.

These new data elements may, in some cases, serve as highly accurate and helpful tools to guide senior leaders' decision making. In other cases, they may be inaccurate and highly misleading. Despite being built from constituent government data, there is rarely any transparent mechanism to determine which is the case. To provide such a mechanism would be to illustrate to the government how the data elements are built, eliminating the need to keep paying the contractor.

These proprietary data elements may ingrain themselves into an office's processes to the extent that the office becomes dependent on the contractor. In these cases, a de facto monopoly arrangement can arise because the contractor is the only company with access to the computer code used to make new data elements out of the government data.

[31] Ibid, 16.
[32] GAO, *Government Is Analyzing Alternatives for Contractor Identification Numbers,* June 12, 2012, accessed July 16, 2018, https://www.gao.gov/assets/600/591551.pdf. The GAO reported that contract costs increased from $1 million per year in 2002 to $19 million per year in 2012.
[33] Ibid, 4.
[34] Ibid.

**Case Study:
Black-Box Data Elements**

In recent years, AI has become a prominent buzzword in Washington, DC, policy circles, leading to increased interest in DoD expenditures on AI-related services and solutions. There is no AI description in the product, service, or industry coding systems used by the government for tracking contract award data. Some policymakers have expressed a desire in recent years for data on government procurement spending that might be considered AI-related.

At least one company provides an answer to this question. The company runs automated analysis on data from FPDS and other sources to determine whether individual transactions are likely to be AI-related or not. This analysis produces aggregate numbers for the whole U.S. government, which are displayed in a dashboard interface that can be filtered and customized by the user.[35]

The resulting data may or may not accurately convey whether a transaction is AI-related, but one problem is that there is no way to know for sure. The data are essentially generated via guesswork. The guesswork is performed via computer code written by well informed people, but it is guesswork nonetheless.

Possible inaccuracy is only one problem involved in this process. The other is that due to the lack of methodological transparency, DoD has no way to generate the new data elements on its own. If senior leaders begin to expect the regular reporting of those data elements, DoD will be dependent on a monopoly vendor to provide it.

There is a way to address both the accuracy problem and the monopoly problem: Have people at the working-level collect and report the desired data elements. It would cost additional time and money, and in some cases, it may require involving non-DoD stakeholders—for instance, GSA agreeing to add new functions to existing IT systems. Some data collection efforts would likely be cost-prohibitive. For this reason, policymakers would need to be willing to engage in honest, good-faith discussions about which types of data collection fail to produce a sufficiently high return on investment.

None of these issues suggest that data analytics firms do not add value to DoD acquisition efficiency and oversight. There is an enormous difference, however, between analyzing data and maintaining key parts of a system's data architecture. A firm that adds value by analyzing data can presumably be replaced with another, more competitive vendor. A firm that owns the sole means of creating certain data elements is effectively a monopoly.

Discussion

The ecosystem of federal acquisition data is enormous, but it suffers from an endemic lack of standardization. This situation leads to inefficient government operations, poor communication, and less effective acquisition systems. There are two ways to address this problem—one easier and shorter-term, the other more difficult but ultimately better in the long run. Recognizing that the best solution may be too expensive and would take too long at this time, a middle approach to improve the sharing of data would offer near-term benefits.

The easy solution is to add translators as needed to convert the data architecture from one IT system into the architecture of another IT system. This approach is currently embraced by DoD and the rest of the federal government. A prominent example is USA Spending, a Treasury-run website that converts several nonstandardized data architectures into a common architecture.

[35] Data analytics firm, presentation to Section 809 Panel, mid-2018.

The upside of this approach is that problems can be addressed on an ad hoc basis, and problem-solvers can build solutions without first obtaining buy-in from senior decision-makers and data system owners. The downside is that it essentially involves plastering additional interfaces on top of legacy systems. This approach adds a new layer of data, but does not actually streamline the underlying architectures of data collection systems. For this reason, it retains the added costs and inefficiencies associated with having too many translators involved in the process.

The more difficult solution would involve consolidating existing data architectures into a single common architecture and getting all stakeholders to agree to its adoption. There would be substantial technical, bureaucratic, and political challenges.

- **Technical difficulties:** No single person knows enough about the governmentwide acquisition and financial data to singlehandedly redesign the entire architecture. A high-quality architectural redesign would have to involve working-level people with data science expertise in requirements, budget planning, appropriations, resource management, contracting, and other fields. To maximize interoperability across functional communities, it would be preferable to adopt existing commercial standards to the maximum extent practicable.

- **Bureaucratic difficulties:** Involving so many stakeholders would create a danger of stagnant committee decision making, universal veto authority, and scope creep. There might also be a danger of *not invented here* syndrome impeding adoption of already-existing data architectures—for instance, those borrowed from National Institute of Standards and Technology, industry organizations, multilateral institutions, and the open-source developer and data science community.

- **Political difficulties:** In many organizations that would be affected by a redesign of defense acquisition data architecture—for instance, congressional committees, OMB, GSA, and private companies—no one reports to the Secretary of Defense. For this reason, it would be critical to obtain advance buy-in from these stakeholders if modifying data architectures with jurisdictional overlap.

Costs and Benefits

Reorganizing portions of DoD's acquisition data architecture would cost substantial time, effort, and money. There are also substantial costs associated with the status quo. A notional breakdown of the major costs associated with the status quo approach would include the following:

- Continued inefficiencies in DoD operations at the working level.

- Poor communication across functional communities due to widespread misunderstandings of the way other offices do business.

- Labor immobility leading to both inflated salaries and a frequent lack of the right type of highly specialized data science expertise.

- Continued investment in software maintenance to ensure continued accuracy of data translation.

Maintenance of legacy IT systems would be one of the most expensive costs associated with maintaining the status quo. Software systems used in acquisition and finance frequently must communicate with one or many other software systems. If the data structure used in one system is modified, other systems must be modified to accommodate the change. Changing one piece of software or business process can in some cases require dozens of changes to other systems—building new translators to allow for intersystem communication. This can generate enormous costs.

It is not possible to calculate with certainty the percentage of IT upgrades dedicated to building new data translation functions into existing systems. GAO analyses suggest the total amount invested annually in IT modernization is close to $75 billion.[36] If only one-tenth of this money went toward building data translation features, costs would stretch into the tens of billions of dollars over just a few years.

Major costs associated with the reorganization of data architectures would include the following:

- The direct dollar costs of staffing and researching the modifications to be made.
- Efficiency and implementation costs of business process reengineering.
- Costs associated with redesigning existing IT systems to be compliant with new standards.

Ideally, uniform data architectures would use existing commercial standards. They would be designed in such a way that preexisting, customized, business software could be replaced with simple software based largely on code in the public domain. If a data standardization effort were to abide by these principles, the costs of such an effort could in many cases be much lower than the costs of maintaining the status quo.

Short-Term Solutions Using Proprietary Interfaces

Many companies now offer services wherein they download bulk federal government data, add customized data elements to it, and sell visualizations using those data elements back to the government. This practice may be an acceptable short-term approach in some cases. If policymakers come to rely on visualization tools that simply collect from preexisting systems, there is a danger that it will gradually degrade the political incentives to reform the underlying data architecture. Without this kind of reform, functional communities will continue to lack architectural alignment and institutional inefficiencies will keep increasing.

Conclusions

If acquisition financial data architecture were to be comprehensively reformed, senior U.S. government leaders would need to collectively commit to standardizing and consolidating existing architectures. The data architectures used by the many functional communities would need to be aligned to portfolios of mission capabilities.

---

[36] Based on GAO's FY 2018 governmentwide IT spending estimate of $96 billion and analysis that the federal government spends "more than 75 percent" of IT investment on operation and maintenance. GAO, *Information Technology: Federal Agencies Need to Address Aging Legacy Systems, GAO-16-696T,* May 25, 2016, accessed August 17, 2018, https://www.gao.gov/assets/680/677454.pdf. GAO, *Information Technology: Further Implementation of Recommendations Is Needed to Better Manage Acquisitions and Operations, GAO-18-460T,* March 14, 2018, accessed August 17, 2018, https://www.gao.gov/assets/700/690655.pdf.

This process would be technically challenging and politically contentious. It would affect the way business is done in congressional committees, OMB, GSA, and other organizations where no one reports to the Secretary of Defense.

Regardless of the challenges, such a reform would have to occur eventually to allow for meaningful comparison of acquisition and financial data analysis across functional communities. Without common architectures, a data analyst from the contracting community will not be able to reliably trace a program back to the budgeting data from which it was funded, let alone the requirements data from which the program originated.

One option is to continue spending billions of dollars building interfaces to translate between IT systems, or providing additional education and training to data analysts. Continuing to speak several different data *languages* across communities, however, will mean data analysis continues to be untrustworthy and inefficient because of the need to include various human and machine translators in the process. Many of the machine translators are obscure IT systems to which virtually no one pays attention until they break or produce obvious errors. This approach leads to unnecessary delay, dysfunction, and maintenance costs. Many of the human translators are the only individuals in their workplaces with certain skill sets, meaning inefficiency in the form of backlogs, wait times, and potential labor market distortions.

A long-term way to address these problems would be for senior-level stakeholders to appoint experts; assign these experts to build a uniform, top-level data architecture; and commit in advance to implementing that new architecture throughout the government's acquisition and financial system, which would likely take decades.

Case Study:
Trade Data Standardization[37]

In the mid-1800s, governments used a patchwork of different categorization systems to apply tariffs and collect statistics on goods being traded internationally. A trader importing American wheat into France, for instance, might have his product defined and assessed using a completely different framework from that applied to the same wheat being imported into Spain. This lack of standardization across jurisdictions led to confusion, excess complexity, and higher transaction costs in the international trade community.

These problems were acute enough that beginning in 1853, a global community of statisticians and economists began to meet regularly to develop a uniform tariff data architecture. Despite holding congresses on the matter every few years, it took decades before participants made even small headway on real-world implementation. Early efforts to create a standardized tariff data architecture included the Austria–Hungary Tariff of 1882 and the 33-country Brussels Nomenclature of 1913.

[37] Historical details on 1882 Austria-Hungary tariff, 1913 Brussels Nomenclature, and 1931 League of Nations tariff data architecture from Howard L. Friedenberg, *The Development of a Uniform International Tariff Nomenclature,* published by U.S. International Trade Commission (then U.S. Tariff Commission), April 1968, accessed September 7, 2018, https://www.usitc.gov/publications/tariff_affairs/pub237.pdf. Other historical information from World Customs Organization, *The Harmonized System: A Universal Language for International Trade,* 2018, accessed September 7, 2018, http://www.wcoomd.org/-/media/wco/public/global/pdf/topics/nomenclature/activities-and-programmes/30-years-hs/hs-compendium-2018_v4_june-2018.pdf.

Case Study:
Trade Data Standardization[37]

In 1931, under the League of Nations, the first widely-accepted tariff data architecture was published. The vast majority of countries today, including the United States, continue to apply tariffs and negotiate trade agreements using a data architecture largely based on that of 1931. Between initial conceptual development and initial implementation, it took nearly a century for the reformed data architecture to come into effect.

The chief lesson of this historical case study is that data architecture standardization can take a very long time. In the case of tariff data, it took decades to reach agreement across jurisdictional communities.

A uniform acquisition financial data architecture would not only stretch across jurisdictional communities; it would also stretch across functional communities—requirements, budgeting, program management, contracting, and others. Both these factors—jurisdiction and functional area—would increase the total time, complexity, and political capital that would be necessary for such an effort to be successful.

In the short term, a more feasible alternative to comprehensively redesigning data architectures would be to iteratively improve DoD data management by improving business processes and eliminating IT interfaces. The CMO would be the obvious authority to lead such an effort. Implementation of DoD data management practices would be largely a technical endeavor, not a policy one. The CMO should be tasked with identifying redundant business processes, process nodes, and IT systems for elimination.

Implementation

Legislative Branch

- Direct DoD to identify and consolidate or eliminate competing data architectures within the defense acquisition and financial systems.

Executive Branch

- There are no regulatory changes required for this recommendation.

Implications for Other Agencies

- There are no cross-agency implications for this recommendation.

THIS PAGE INTENTIONALLY LEFT BLANK