WHITEPAPER

Peraton

BEST PRACTICES FOR DATA MANAGEMENT WITHIN DEFENSE ENVIRONMENTS

DO THE
CAN'T BE DONE.

EXECUTIVE SUMMARY

There is ongoing debate about the ideal data management framework for the U.S. Department of Defense (DoD) to respond to emerging challenges from peer competitors and related global cyber threat actors. Without a framework consensus, rapid and non-standardized data capability adoption may sometimes occur to respond to burgeoning threats.

This whitepaper expounds upon concepts first presented at the 2023 AFCEA TechNet Innovation Summit in Augusta, GA. Specifically, this whitepaper discusses mainstream DoD data management and technical execution challenges, proposed data management practice enhancements to potentially overcome them, and finally a Peraton recommended data management framework with proven effective data management best practices for the present and future.

DATA MANAGEMENT CHALLENGES

Currently, there are numerous challenges to implementing successful data management within the DoD. Many of these data management challenges to improve mission effectiveness can be attributed to wider (non-DoD) data assumptions, versus any adversary driven problem. These assumptions may include the following:

- **Existing (Data and Practice) Workflows:** DoD end users produce, consider, assess, and respond to data in very specific ways significantly shaped by mission requirements. Still, most DoD data management adoptions rarely consider how end users interact with their data to accomplish a specific mission, nor the workflows they undertake to do so. This can lead to underutilized capability or technology abandonment when data access is too unfamiliar and/or cumbersome.
- **Data Nuance, Structure and Automation:** When data integration is a goal in a data management framework, insufficient attention is sometimes granted to data nuance, differences in data structuring, or how/when to automate data best (i.e., for data treatments). While ignoring these areas is a minor problem with smaller data sets, it can become a significant challenge, if not impossible task, when factoring in petabytes of data amongst dozens of sources. In these latter occurrences, data can become isolated, maybe visualized, but – without finesse applied against it - almost never combined into a greater strategic picture. Once more, technology abandonment or expensive integration costs (to close gaps) can become unwanted outcomes.
- **Human: Machine Teaming Efficiencies:** The DoD is not alone in ascertaining how to best rack-and-stack human and machine (Artificial Intelligence/Machine Learning, 'AI/ML') analysis within an assessment workflow. This can unintentionally lead to humans having to perform laboriously doublecheck automation against errors at unnecessary stages as that process step is better suited for more contextual, manual versus automated data engagement. Another unintended consequence is underutilized automation steps (i.e., large scale, simple data cleaning and/or data integration prior to focused contextual looks) otherwise capable of yielding the necessary process efficiencies, which are especially important when time is mission critical.
- **Appropriate AI/ML Model Design Validation:** Enthusiasm abounds within the greater U.S. Government (USG) to locate and implement automated data technologies, although lesser consideration accompanies it on how to best care, nurture and feed its models. Any AI/ML model must undertake continuous testing and validation to ensure it remains accurate while also compatible with future data sources and capabilities. In concert, new and relevant AI/ML models should be proactively developed to stay ahead of emerging threats.
- **System Resource Requirements:** As data access and storage pools grow, so, too, does processing and warehousing requirements. While hybrid cloud/multi-cloud solutions are terrific ways of uniting otherwise disconnected environments, failing to account for scalability beyond acknowledging the term (by throwing more processing power into the fray) will produce underperformance (choke points) or even non-performance (disconnects). Therefore, expected data resource requirements should be quantified and planned.
- **User-Centered Design:** Most DoD data-centric capabilities do not specifically execute a comprehensive assessment of how proposed end-users interact with a mission-relevant data management system. This might lead to a current/future user interface and experience that is anything but organic to end users. A good DoD user-centered design could include menu items reflective of mission steps; logically layered visualizations; quick access to deeper human analyses; automated import/export into sister tools; and, finally, ensuring overall look and feel match organization conventions.
- **Historical Data Needs and Archival:** As many USG agencies now require consideration of archival data in assessments, this can produce exorbitant storage, processing, and overall resource costs over time. If current requirements mandate the use of archival data, these costs need to be factored in and stated as part of any holistic data approach, to include expectations they will only grow as more data is introduced then stored.

INTEGRATION AND THE COMMON OPERATING PICTURE

The DoD remains a forward-thinking department. Thus, several steps were taken to overcome the above stated challenges. Some of the DoD agencies have executed capability audits, concluding it best to maximize existing resources in a unified environment when improving, augmenting, or replacing any data management capability.

To many within the DoD, the ideal data management framework culminates in a common operating picture (COP) featuring holistic mission data visualization and supporting assessments. For these COP's (and due to current offerings), co-located capability is typically squeezed into a single, often bespoke, platform comprised of a series of application programming interfaces (APIs) with disparate entry/exit points, separate analyses, and each possessing unique integration requirements. This COP approach can prove costly, as maintaining the capabilities means treating each as separate entities, requiring a plethora of data licenses to feed individual API human integrators and ensure connective tissue to any central platform.

Similarly, cohabitating mission-relevant information does not guarantee improvements to mission effectiveness if they are not comingled. For example, data scrapes of social media, information networks, and geographical intelligence visuals may yield ongoing snapshots of a target or population of interest. However, analyses of such data rarely offer confident measures of effectiveness against an adversary as limited to correlation or time series looks. When this reality emerges, more data or tools are typically purchased -- thinking them proverbial missing pieces with the thought to leave no stone unturned if funding is available to dig wider and deeper. Instead, this well-intentioned effort often exacerbates the problem of abundant yet isolated capability and cost, while increasing human analyst lift to sift through even more information produced from the overall solution.

Even in the best COP approach, the often diverse, large scale and context heavy data powering it is almost never designed to comingle, even for similar mission types and customers. When some COP data integration occurs, in a data fabric or otherwise, analysis options are limited and rarely automated.

At present, most frequent default COP assessments still appear as correlations, data presentations (amounts/weights by time series), and co-association (what items tend to appear with others—and when), all nonsufficient as measures of effectiveness or structured in a fashion where advanced analyses is possible (i.e., via complex machine learning-led predictive modeling) to positively impact mission effectiveness.

In these scenarios, and with limited assessment potential, AI/ML contributions will be minimal. With automated data practices still so new to the DoD, there remains a logical shortage of security cleared DoD data science professionals who simultaneously understand mission requirements and advanced data techniques to potentially build a better COP, thus capable of creating and improving data integration approach to get there.

Within the DoD, a promising opportunity for an improved data COP and data management approach resides in the information warfare space, where efficient and effective understanding of adversary impact on the information environment is needed in real time. At present, however, almost every information warfare data assessment capability does not formally account for entropy—natural state of information environment data 'chaos'—or the pattern of life and how seemingly disconnected phenomenon and the variables which represent them appear alongside each other. Without these considerations, effective measures for information operations may be ripe with Type 1 or Type 2 error, false positives, or negatives, respectfully. Once more, this information environment assessment hurdle is certainly not limited to the DoD alone.

Finally, with so many new data management capabilities and models in the DoD ecosystem, conventions, standardization, and uniformity are a challenge. Those responsible for maintaining the security of these systems possess the most difficult task of all, defaulting to most conservative approaches for data handling to protect against spillages and classification. Specifically, new data management techniques and methods, including data integration, semi- or automated data cleaning, joint visualizations, and hybrid cloud multi cloud environments, can lead to security violations by default if not properly accounted for and controlled.

KNOWN PRACTICE CHALLENGES AND PERATON RECOMMENDED SOLUTIONS

With the right mindset, all challenges are quite solvable: this is an opportunity not a capability issue. Importantly, adversary advantage does not stem from outthinking, outspending, or outmaneuvering the DoD. Below are some current DOD practice issues and recommendations on remedies.

Overtly Identify and Embrace the Challenge

Challenge

In some instances, shortcomings in DoD data management frameworks reside in disconnect between the actual problem and the role of data in solving it. For example, peer competitor misdeeds in Europe in Southeast Asia may seem so daunting and amorphous that there are no means of confidently scoping such large problems, let alone to locate and deploy sufficient capability to stay ahead of their actions.

Recommended Solution

Recognize the challenge for what is—in tangible, smaller bites—then begin to thoughtfully apply data capability to address it from these lenses.

Use/Re-use All Data by Default

Challenge

Similarly – when dealing with peer competitor challenges - purchasing additional data appears logical to produce an increased (and, ideally, holistic) strategic perspective. Still, strategic perspective rarely stems from what information is at one's disposal, but rather how such data is considered. This could include using the same data in a different fashion, increased awareness and assessment achieved from data integration, automated data structuring techniques, proactive threat identification from predictive modeling, etc.

Recommended Solution

Before purchasing additional data sources, conduct a thorough audit of all existing, ensuring all current and historical information is considered and, in all manners, possible.

Increase Data Confidence

Challenge

Underwhelming data confidence doesn't typically stem from poor program design but rather its data validation approach.

Recommended Solution

Deliberate and more thoughtful consideration on how measures of effectiveness are constructed, as the confidence levels attached to them are attained and verified. This includes consideration of external validity, where data methodologies incorporated are not biased by operators, are content agnostic, and potentially applicable to other mission situations.

Align Data Resourcing and Solutioning

Challenge

While integration is the goal, most of the DoD data environments, systems and solutions are purchased and maintained separately, and can unintentionally produce disjointed costing, analyses, and workflows.

Recommended Solution

All existing and proposed DoD management framework offerings considered in unison and fit to each other. By default, this reduces future costs, ensures maximum process efficiencies, prioritizes capability, and identifies air gaps needing immediate response.

Embrace Familiar UI/UX

Challenge

A lot of tech abandonment can be caused by too many steps, logins, and/or interfaces unfamiliar to those using them. Ensuring the right people have the right data—and ways of access—can go a long way.

Recommended Solution

Prior to adopting solution alternatives, DoD should first assess how end-users conduct everyday business in current operations then realign, retest, and rethink.

Integrate then Sunset Existing Offerings

Challenge

To some, sunsetting technology equates to sunk costs or mistakes.

Recommended Solution

If legacy capability, data, and processes are thoughtfully integrated into new offerings then logically phased out, the most relevant information, user-centered design, and lessons learned transition with it. Moreover, this alternative approach ensures consideration of historical data in future assessments and model building.

Train, Educate, and Certify Often

Challenge

There is no shortage of training, education, or certification options for DoD technically oriented professionals. Still, many of such offerings are not purposefully aligned with current and future capability requirements and updated by default.

Recommended Solution

Implement training, education, and certification programs reflective of actual mission requirements, and encourage, incentivize, and assist those who work with data science-oriented individuals to upskill.

7 STEPS TO ACHIEVE BETTER DATA MANAGEMENT

Practice and process unity plus maximum user and data access are pre-requisites toward any best practice data management framework. What follows are seven best practice steps the DoD can consider to improve its data practices and related mission effectiveness advised by them.

1. **Adopt a unified security structure with known user types and groups.** At present and within USG, different agencies and partners – each with distinct security classifications and accompanying information access inherent – are expected to contribute to an overall mission. Still, and in the interest of security, many collaborative environments feature an abundance of completely disparate systems to protect against spillage. An alternative data approach is to have a clearly specified user types -- each with similarly specified data access, analysis, and visualizations options -- where any and all actions consider information the user has both a need and right to know. In tandem, establish a single sign-on within a central platform. Combined, this significantly reduces security management and monitoring requirements (as now defined by user type), maintenance, integration, and related technology costs in a more streamlined offering. This approach also aligns with Zero Trust adherence, by reducing the number of vulnerabilities for an otherwise array of non-named users and their island systems.
2. **Work with Hybrid Cloud Multi-Cloud (HCMC) providers.** As a means of uniting a cornucopia of established cloud and on-premises environments across multiple classification levels), limited actual consideration is given to employing this approach within larger organizations (to improve collaboration and data/capability access). Working alongside HCMC providers, DoD can be better served re-examining the potential of them to maximize access, analysis, and mission effectiveness.
3. **Increase consideration of AI/ML models native to a central data platform and tailored to them,** both tested and updated, by default, by the platform provider (accredited by the DoD in that environment). Nearly every major data platform provider now features a series of diverse AI/ML model types within that can be subsequently adjusted to fit most mission or organization needs.
4. **Understand the real value of an API ultimately lies in the data it considers.** By rethinking APIs, not as a separate capability rather by contribution of its data, this may greatly expand possibilities to create better COPs with more robust analysis potential (via data integration). To do so, some API contributions to a COP may be better considered as data imports in a single, combined data picture easily accessible by analysts versus a separate application where capability redundancy is a distinct and expensive possibility. This is a paradigm shift; thus, it may prove helpful for the DoD to have candid discussions with application providers on alternative ways of maximizing their offerings together.
5. **Realize reducing APIs as standalone applications also reduces the number of integrators needed to ensure operability**, plus the possibility of system disconnects and/or stalls when new data and capabilities introduced. If this alternative is adopted, ponder – in unison - a rebalance of integrators to maintain a more centralized data management system than an array of applications (formerly in isolation). Ensure data provenance throughout. As historical data more prominently required and considered in AI/ML-powered data assessment approaches, data provenance becomes essential to data confidence also if any proactive, predictive modeling (via ML) expected.
6. **Ensure data provenance throughout.** As historical data more prominently required and considered in AI/ML-powered data assessment approaches, data provenance becomes essential to data confidence also if any proactive, predictive modeling (via ML) expected.
7. **Create and execute technology roadmaps with direct linkages and justifications to mission requirements.** Discuss and capture most realistic future mission needs, and roadmap supporting technology to get there ahead of time and via logical, incremental growth.

CONCLUSION

As always, the best data management framework is 100% mission-centric. It recognizes that data questions and outputs are not technical issues solely for a J6 organization to tackle. An effective data management framework facilitates how the overall mission is planned, executed, documented, and assessed. This must be accompanied by building and sustaining a core user base of data advocates across an organization, individuals collectively committed to continuously ensuring solutions are truly meeting expectations while holding deficiencies accountable. Data management frameworks are only successful when the organization leadership ensures the technology providers, strategic integrators, and all contributors to such efforts are equally contributing. Importantly: never adopt or abandon any part of a data management framework without sufficient knowledge as to why.

Remaining vendor-agnostic is therefore critical. Becoming fiercely loyal to any single data management solution or data provider limits access to expanded data management best practices. It's best to embrace a research and development mindset and stay committed to building and testing AI/ML models on a continuous basis. Ensuring all data management framework features sufficient external validity to support beyond today's problem.

Finally, be good stewards. Surges are a part of any military conflict, and increased technology investment will occur alongside them. Always look for efficiencies while doing so, as dramatic capability growth without a proper playbook will create chaos, confusion, and redundancy. Training and education that sticks is essential: create and nurture a future leaning, non-siloed workforce to champion data driven approaches.

HOW PERATON CAN HELP

Peraton possesses a mature and available data management framework capable of guiding DoD customers to achieve truly data centric operations. This data management framework draws upon our experience building and modernizing the data management solution for one of the largest data sets in the intelligence community. Toward this end, and as evidenced by sentiments expressed in this whitepaper, Peraton's data management framework focuses upon mission needs while increasing efficiencies and effectiveness along the way toward maximum effect. Peraton's data management framework ensures an organization's culture, processes, and technology investments are maintained, and matches outcomes to fit its strategic vision, risk profile and budget. Lastly, Peraton's extensive commitment to research and development through Peraton Labs, its internal research and development portfolio, and continuous innovation simultaneously enable ongoing discovery and technological agility.

Peraton

ABOUT PERATON

Peraton is a next-generation national security company that drives missions of consequence spanning the globe and extending to the farthest reaches of the galaxy. As the world's leading mission capability integrator and transformative enterprise IT provider, we deliver trusted, highly differentiated solutions and technologies to protect our nation and allies from threats across the digital and physical domains. Peraton supports every branch of the U.S. Armed Forces, and we serve as a valued partner to essential government agencies that sustain our way of life. Every day, our employees do the can't be done by solving the most daunting challenges facing our customers. Visit Peraton.com to learn how we're safeguarding your peace of mind.

Scan to learn more at
peraton.com/capabilities/cyber/