Joint Data Mesh – A Data-Centric Approach for Modeling & Simulations

Samuel Chambers	Jay Freeman	Walter Cedeño	Colby McAlexander
Joint Staff J7	CAE USA	CAE USA	CAE USA
Suffolk, VA	Orlando, FL	Orlando, FL	Arlington, TX
samuel.d.chambers@mail.mil	jay.freeman@caemilusa.com	walter.cedeno@caemilusa.com	colby.mcalexander@caemilusa.com

ABSTRACT

The Joint Force needs a reliable, realistic, relevant, repeatable, and recordable training environment to rapidly model, simulate, experiment, rehearse, and inform operational plans, yet it lacks an optimized data-centric architecture balancing trust, security, and data access at scale to improve readiness reporting, analytics, AI/ML development, and decision-making. While the Joint Live Virtual Constructive (JLVC) federation has evolved for 20+ years, modernization is required to meet future training and enable faster integration and data-sharing across Service, agency, and partner simulations. Currently, the Joint Event Lifecycle (JELC), from initial planning to execution with simulation, is nominally 12-18 months. Historically, very little data generated during a JELC was retained in a visible, accessible manner, thus, losing the opportunity to capitalize on this rich dataset for meaningful post-event analysis, machine learning, and advanced data analytics. This paper demonstrates how employing data mesh concepts, the Joint Training Tool (JTT), and existing JLVC applications, will address these issues. There are three aspects to our proposed solution. First, modernization should incorporate proven governance processes and standards. Second, we illuminate a 'data-federation' framework that encourages a 'data-centric unity of effort' that is founded on these core principles: domain-oriented, decentralized data ownership; data-as-a-product; reduced total cost of ownership; orchestration and automation to respond to changes. A self-service infrastructure enhanced by an ontology framework promotes data consistency, quality, fitness, and relationship linkages. Finally, established on a Zero-Trust Architecture (ZTA) and scalable computing platform, we highlight architecture components such as registry Application Programming Interfaces (API) to ease data registration/exploration; read semantic layer schema to provide standardized data views; data access layer to enable domain agnostic data retrieval. Our approach delivers agility and flexibility to quickly consume authoritative data at speed and scale, ready for today's demands and the future proliferation of data sources and consumer-driven analytics.

ABOUT THE AUTHORS

Samuel D Chambers is a Scientist with the Environment Operations Division (EOD) of the Deputy Directorate Joint Training Joint Staff J7 in Suffolk, Virginia. He manages the development of the Joint Training Tool (JTT) and Joint Live Virtual Constructive (JLVC) federation data services and serves as the Data Steward for the Joint Staff J7. In this capacity, he is a co-lead for the data line of effort (LOE) for the nascent JLVC modernization program.

Jay Freeman is a Synthetic Environment Technical Fellow at CAE USA. His project oversight includes: decision support R&D investments; USSOCOM Mission Command System and Common Operating Picture; Joint Staff J7 EOD's Joint Training Tool development. Prior employment: System/Software Architect for SE Core DVED, TERREX, Lockheed Martin STS (ATARS - SOFPREP) and Intergraph Services Company. B.S. from Hobart College (Geneva, NY); M.S. from University of Alabama in Huntsville (Huntsville, AL).

Dr. Walter Cedeño is a Lead Data Scientist at CAE with 25-years of experience in: AI/ML; Operations Research; HPC; Soft Computing; Big Data; Data Engineering. B.S. and M.S. in Applied Math, University of Puerto Rico, Rio Piedras; PhD in Computer Science, University of California, Davis. Prior: Adjunct Professor, Penn State Great Valley; Principal Data Scientist, Johnson & Johnson; Software Engineer, HP/Agilent, LLNL.

Colby McAlexander is an Advanced Concepts Engineer at CAE USA. Retired Naval Aviator, Information Warfare Officer (25-years). Experienced in joint, multi-national, and all-domain operations, plans, exercises, and wargames; lead IW expert in the highest multi-classification, pedigreed multi-threat war games at U.S. Naval War College. B.S. in Computer Science, Texas A&M University, College Station, TX; MOAS, Air University, Montgomery, AL.

Joint Data Mesh – A Data-Centric Approach for Modeling & Simulations

Samuel Chambers	Jay Freeman	Walter Cedeño	Colby McAlexander
Joint Staff J7	CAE USA	CAE USA	CAE USA
Suffolk, VA	Orlando, FL	Orlando, FL	Arlington, TX
samuel.d.chambers@mail.mil	jay.freeman@caemilusa.com	walter.cedeno@caemilusa.com	colby.mcalexander@caemilusa.com

INTRODUCTION

The Joint Live Virtual Constructive (JLVC) Federation was established and has evolved for 20+ years to meet joint combatant command (CCMD), service, and agency joint training needs; It is comprised of 34 tools, integrated by the Joint Staff (JS) J7, and is now modernizing to enable faster integration of simulations. Recent JS J7 investments have focused on data consistency to support the JLVC Federation and include:

- ❑ The development of a Joint Training Tool (JTT)—a web application that acts as a one-stop shop for the machine-readable planning & design of a training event;
- ❑ Force structure and terrain data to directly link planning data to Modeling & Simulations (M&S).

The JLVC Modernization Project is a 5-year plan and will incorporate Cyber and Space Domains, as well as the Electromagnetic Spectrum capabilities to meet the guidance of the 2022 National Defense Strategy (NDS) and the Joint Warfighting Concepts (JWC) to defend the homeland and our allies against multi-domain threats. A key JLVC modernization line of effort (LOE) is transformation to a data-centric architecture and 'analytics ready' federated organization – aligned to Combined Joint All Domain Command & Control (CJADC2) – to meet the NDS and 2020 Department of Defense DoD (DoD) Data Strategy (DoDDS) visions. This LOE focuses on the implementation of the DoDDS framework (Figure 1), to include making data visible, accessible, understood, linked, trusted, interoperable, and secure (VAULTIS). DoD writ large understands how to acquire, train, maintain (E.g. systems, unit, phase, and depot level), lifecycle, and modernize its forces, weapon systems, ships, armored personnel carriers, aircraft, satellites, cyber effects, etc; however, the same is not consistently true with our data. Yet as DoDDS states, "Data is ubiquitous…Data underpins digital modernization and is increasingly the fuel of every DoD process, algorithm, and weapon system." Both the NDS and DoDDS emphasize data's criticality:

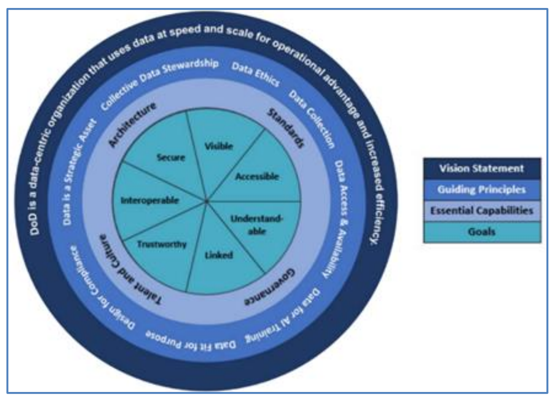

Figure 1. DoD Data Strategy

- ❑ NDS— "The [DoD] is establishing a new framework for strategic readiness, enabling a more comprehensive, data-driven assessment and reporting of readiness to ensure greater alignment with NDS priorities."
- ❑ DoD Data Strategy —"Data is a Strategic Asset – DoD data is a high-interest commodity and must be leveraged in a way that brings both immediate and lasting military advantage." "DoD must accelerate its progress towards becoming a data-centric organization." Specifically, to enable data-driven, insights and calculated risk decisions that allow "execution of swift and appropriate action."

The JLVC Federation's realization of these requirements for optimized joint training is through modernized data stewardship processes (i.e. Governance) and federated data-mesh architectures, with associated decision-support systems, to effectively discover, process, evaluate, integrate, and visualize data at speed and scale. This paper's proposed concepts and solutions align with JLVC modernization efforts and an 'analytics-ready data' transformation for measured and cohesive joint force development, arrangement, employment, and sustainment against priority strategic challenges and threats. Furthermore, this paper highlights mechanisms to:

- ❑ Improve data consistency, coherency, timeliness, and effectiveness for JS J7 data-driven decisions;
- ❑ enable cross-domain solutions (CDS), mission partner environment (MPE), and zero-trust architecture (ZTA)
- ❑ Optimize data-sharing at "the speed of the fight" across the joint force and mission partners for decision advantage across the continuums of training, competition, and conflict.

Our proposed Joint Data Mesh (JDM) is based on the data mesh concepts introduced by Zhamak Dehghani in 2019; It delivers an environment for managing a diverse set of data sources, shaped to meet DoD, JS, CCMD, Service, Agency, and mission partner data requirements. Leveraging VAULTIS and ZTA principles, the JDM decentralizes data management to groups of teams that employ common industry/interservice best practices and governance provisions; The JDM delivers a data-centric and diverse consumer-based oriented solution to best support data landscape changes, proliferation of data sources, and consumers, stakeholder analytics needs, and agility to respond to change.

BACKGROUND

A CCMD-level event supported by the JLVC Federation nominally requires a 12–18-month Joint Event Lifecycle (JELC) process to plan, design, and execute. Historically, planning was performed with analog products (Microsoft applications) stored in an event-specific SharePoint portal, which is not fully efficient for data linking and sharing.

Joint Training Challenge

Aforenoted analog products require simulation data specialists to develop or modify existing simulation data to support the Master Scenario Event List (MSEL) injected into the training audience at a specified time. Combatant Commands support this design to closely manage the operational and strategic injects for their events. In this way, exercises are labeled as 'MSEL-driven, simulation supported'. If a set of JLVC simulations are not capable of MSEL modeling, a manually written 'white card' is used to initiate the actions of the training audience. Additionally, deployable training teams (DTT) and military subject matter experts (SME) exercise designers often have little-to-no knowledge of M&S applications' capabilities. They often lack an understanding of how data is stored, maintained, translated between systems, and modified to meet training objectives. White cards too often fill the gap between this lack of understanding of operational requirements and knowledge of available data and software functionality. Past attempts to create a data-centric synthetic environment replicating an operational or information environment failed for a variety of reasons, including:

- ❑ a centralized approach of collecting data in a central repository, managed by a single organization is not sustainable, as the volume of data required and the separation of data from the domain expertise is not ideal;
- ❑ the immaturity of Application Programming Interfaces (API) in DoD infrastructures;
- ❑ no overarching guidance and cultural challenges, until the 2020 DoDDS directing the DoD to become a data-centric organization.

Early Game Changers

To enable Joint Training, the JS J7 designed and developed the Joint Training Information Management System (JTIMS). JTIMS was created as a web application to enable exercise planners to plan and design, an exercise, and assess events. One of the capabilities of JTIMS was the ability to manage MSELs in a single location. JTIMS was connected to other systems and was able to pull in Combatant Command Joint Mission Essential Tasks, derived from the Universal Joint Task List (UJTL). The UJTL Task Number is a key field that enables linkages across multiple applications and data. Coupling the UJTL Task Number with other key fields (e.g., Unit Identification Codes (UICs); Distributed Interactive Simulation (DIS) enumerations, etc.) enables performing advanced queries on joint training data to derive context and information on Joint Force readiness. Thus, the UJTL is the common thread between joint processes and systems. One can follow a specific task through an organization's readiness, training, lessons learned, and back to readiness. Gibeling (2022) provides a notional example (Figure 2): CENTCOM determines 'UJT ST 5.4.4 Coordinate Security Cooperation' task is essential to their Operation Inherent Resolve (OIR) mission. They assign ST 5.4.4 as a part of their Joint Mission Essential Task List (JMETL). When planning for the exercise titled 'Internal Look', CENTCOM builds training objectives for ST 5.4.4 in JTIMS. These training objectives may be supported by

JLVC simulations. The vignette uses military units in the Order of Battle Service (OBS) format, including UICs for units and DIS enumeration for entities. During the exercise, observations and Training Performance Assessments (TPAs) are captured on ST 5.4.4. The TPAs are used to link training information to readiness posture.

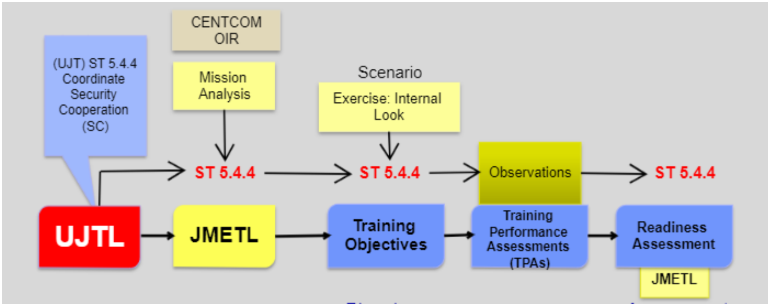

Figure 2. Flow of a UJT through the Joint Training System

JLVC Modernization

The next evolution for JLVC was transitioning from JTIMS to the JTT to integrate training data generated during exercise planning with simulation data, moving towards a data-centric environment. The Joint Staff intends to modernize and transition the joint components of JLVC into the JTT architecture. The JTT captures, processes, and provides data necessary for all aspects of joint training; It uses modern web and data exchange technology to move beyond stove-piped application-centric systems tightly coupled to proprietary data formats. Moreover, JTT enables the identification of training requirements, development of training plans, exercise force support requests, and the design, execution, assessment, and reporting of exercise(s) results to increase readiness across the Joint Force. Data sharing between the JTT and multiple systems is facilitated by enterprise APIs and the employment of a well-documented data model and appropriate meta-tagging of data.

The continued modernization of JTT will both enable the use of current technologies and facilitate an evolutionary transition from application-centric applications and simulations to a data-centric, web-based, and single digital environment. This approach maximizes commercial web-based technology, open-source software, and recognized technical standards which enable compliance with the DoDDS and the DoD Creating Data Advantage Memorandum's guidance (i.e., "use automated data interfaces that are externally accessible and machine-readable; ensure interfaces use industry-standard, non-proprietary, preferably open-source, technologies, protocols, and payloads."). This modernization is a perfect opportunity to develop and implement a new JS J7 Data Governance Framework and JDM to meet DoDDS directives to develop a measurable Data Strategy Implementation Plan.

DATA GOVERNANCE

Data is directly proportional to technology advancement and inverse to cost – as technology advances and costs go down, data volume increases. DoD has a distinct exponential growth in the volume of data created, processed, analyzed, stored, transmitted, and shared. Data without governance can become like the "wild-west," and inhibit the decision-making it should enhance if it is no longer valid. Governance is a positive mechanism to deliver value across one or more organizations, reducing costs, minimizing errors, facilitating shared understanding/knowledge, ensuring compliance with policies and law, and maintaining the security and integrity of the data. Our overall strategy and prototyped architecture we present in this paper consider an effective data governance framework that promotes data that is fit-for-a-purpose and discoverable, with associated best-practice ontology and provenance paradigms; The target result is "the Department's ability to share information, fight if necessary, and win conflicts in an era of great power competition, and it will enable operators and military decision-makers to harness data to capitalize on strategic and tactical opportunities that are currently unavailable" (DoDDS). We emphasize that an architecture built without effective governance can lead to data and decision stagnation and increased costs. As part of modernization, the JLVC Federations requires a foundational Data Governance Framework which:

- ❑ incorporates metrics, key performance indicators (KPIs), and dashboards to support decision-making, training plans, readiness, and data-centric technology architecture.
- ❑ includes people, processes, activities, and technology.
- ❑ sustainable 'analytics-ready data' transformation for measured and cohesive joint force development, arrangement, employment, and sustainment against priority strategic challenges and threats.
- ❑ synchronized, coordinated, and integrated activities (i.e. unified action) across data owners, data stewards, and internal and external consumers to achieve cooperation toward common objectives (i.e. unity of effort).

Finally, a JS J7 modern Data Governance Framework, coordinated with the JLVC Federation, should enable the decentralized use of data, by provisioning services and tools, and defining semantics and standards for sharing federated domains' data (i.e., an interoperable method to institutionalize, organize, and operationalize data regardless of where the data is stored and managed). The goal is to create repeatable, testable, trusted data processes and activities directly supporting operational, training, and technology priorities. Priorities for both a data governance framework and a JDM include: MPE; wargaming design; data standardization; operational information environment (IE); effects integration; capabilities enhancements; and operational and logistical sustainment in contested environments; enabling enterprise data as a strategic asset.

Data governance includes people, processes, activities, and technology; A data governance framework enables repeatable, testable, trusted data processes and activities to achieve comprehensive, data-driven assessments and reporting of readiness.

ARCHITECTURE DESIGN ELEMENT TO ENABLE ENTERPRISE DATA AS A STRATEGIC ASSET

We aim to facilitate the modernization, use, and sharing of JLVC Federation's data for analytics, awareness, and decision-making by internal and external consumers through the continuance of a synchronized 'data federation' methodology vs. a 'data consolidation' cultural mindset. A 'data federation' aligns to and promotes a centralized data governance framework–with data owner representation–and decentralized data architecture, management, and execution at the point of origin/ownership, complying with Figure 1. While 'data consolidation' presents issues such as:

- ❑ Increased total cost of ownership due to data duplication and complexity;
- ❑ increased risks of data staleness, consistency, security, reliability, and creation of a single point of failure;
- ❑ disadvantageously distances data from the source (i.e. hastens the decrease of data's understandability, trustworthiness, and fitness-for-a-purpose – to name a few).

Moreover, a 'data federation' encourages a 'Data-Centric Unity of Effort'; Multiple distributed data-centric architectures/databases functioning as one—synchronized, coordinated, and integrated. In other words, the concept is harmonized efforts across multiple organizations and technologies advancing towards similar objectives. A 'data federation' approach facilitates on-the-fly modernization and MPE integration at speed and scale. A Data Mesh, from a modernized technology standpoint, champions unified action and unity of effort across all mission partners.

Evolution of Data Architectures

Centralized data architectures have existed since circa 1960s when data warehouse architecture played an important role in extracting data from operational databases and sources to provide descriptive analytics information. Data warehouse key aspects include the use of a universal schema; the design of reference information using dimension tables; the design of fact tables to provide metrics and measures; and support for fast SQL queries using well-defined indexes. However, enterprise-level data warehouses' complexity (e.g. hundreds of extract, transform, load (ETL) pipelines), plus their lack of scalability and flexibility, led to the introduction of data lakes circa 2010. Data lakes addressed these challenges by simplifying ETL to simple copies of the raw data in a centralized repository. Data Scientists and Analysts using the data would transform and clean the data as needed for their machine learning (ML) and business intelligence applications; however, duplication of work and different interpretations of the data would often result when different teams used the data. Also, flexibility and scalability improved by storing the data in its original form in object stores using popular storage formats, but at the cost of losing fast data access SQL provides. This led to data lake houses which combine the data warehouse and data lake architectures to incorporate the advantages of each architecture. Yet, all three of these approaches centralize data in a common environment, require highly specialized data engineers, and drive the data architecture and organization into a monolithic approach.

The introduction of the data fabric concept in the mid-2000s provides a means to connect disparate collections of data tools for big data management in a cohesive and self-service manner. Also, a data fabric delivers capabilities to simplify data access, discovery, transformation, integration, security, governance, lineage, and orchestration. A data fabric provides data management for environments that include operational data stores, data warehouses, data lakes, and data lake houses. Although data fabric storage could be decentralized, the data management and tools are amalgamated and administered by a centralized team. Thereby, a data fabric incurs the issues noted above regarding 'data-consolidation data consolidation.' This led to a data mesh, which presses towards a bottoms-up way to manage organizational data; It empowers businesses to build domain-specific systems to meet their needs while using a common governance framework to support enterprise-data distribution. Key differences between a data mesh and a data fabric:

- ❑ A data mesh is an API-driven solution; A data fabric requires writing code for APIs to interface.
- ❑ A data mesh focuses on people and processes; A data fabric is product/technology-centric (or dependent).
- ❑ A data mesh brings data closer to consumers. A data fabric pulls data away from its source and consumer.

A Data Mesh

Zhamak Dhegihan (2022) defines a data mesh as "a decentralized sociotechnical approach to share, access, and manage analytical data in complex and large-scale environments—within or across organizations … a sociotechnical paradigm [is] an approach [which] recognizes the interactions between people and the technical architecture and solutions in a complex organization." This data management approach optimizes data analytics, data sharing, and user experience for all consumers, internal and external (E.g. decision-makers, data stewards, owners, etc.). Dhegihan (2022) describes four principles to underpin "data mesh's logical architecture and operating model:"

- ❑ **Domain Ownership—**Decentralize data ownership to domains nearest to data (I.e. source; primary users). Enables data optimization, agility, trustworthiness, and resiliency.
- ❑ **Data as a Product (DaaP)—**Data is shared, not consolidated, directly with consumers. Removes **silos,** drives data innovation and lifecycle, and adheres to DoDDS Framework.
- ❑ **Self-Serve Data Platform—**Empowers cross-organization/-functional information sharing. Removes data-sharing friction, streamlines the user experience, and reduces the total cost of ownership.
- ❑ **Federated Computational Governance—**" …creates an incentive and accountability structure that balances the autonomy and agility of domains, with the global interoperability."

Importantly, a data mesh brings the ability to connect to data from diverse data sources (e.g. transactional databases; data warehouses; data lakes; data lake houses; raw data files; and APIs that form an organization's data ecosystem). In turn, a data mesh allows organizations to leverage prior investments while reducing the risk and timeline to adopt a data mesh approach. A data mesh's primary focus is the data itself, not data storage; it best delivers data at speed and scale, even in complex, dynamic, and large-scale OEs. Dehghani (2022) highlights additional beneficial outcomes of a data mesh:

- ❑ "…higher-order value from aggregation and correlation of independent yet interoperable data products;"
- ❑ "Respond gracefully to change: a business's essential complexity, volatility, and uncertainty;"
- ❑ "Sustains agility in the face of growth;"
- ❑ "Increases the ratio of value from data to investment."

A Joint Data Mesh and Joint Data Federation Solutions Enabling Strategic Data-Analytics

Our federated JDM architecture approach (Figure 3) with consumer-oriented APIs enables:

- ❑ consistent use of data across multiple domains;
- ❑ oriented consumer-oriented APIs for data analytics in an agile and minimally obtrusive manner;
- ❑ User interfaces (internal and external) and tools for data access, data exploration, and registry management;
- ❑ access to consumer-specified data, tools, or platforms (e.g., Advana, business intelligence visualizations, etc.).

JDM's foundation is based on an underlying platform that implements a ZTA with Authorization and Authentication in a computing environment and seamlessly scales as necessary for computation increases with the availability of data. The proposed architecture reduces costs, leverages orchestration and automation, and is designed for change, scale, and delivering value to stakeholders via dashboards in a simple, straightforward manner. To note, our cost evaluation measures include the acquisition price, operational execution, and life-cycle costs of technology identified, evaluated, and integrated. As data is a strategic asset, we consider costs across the data's lifetime, such as:

data analytics technology or services; data at rest (server and storage); data access and retrieval (network and APIs); data movement and transformation (extract, transform, and load); data duplication across multiple locations (e.g., on-premise, cloud, staging areas, etc) risks. Orchestration and automation simplify the registration of new data, access to the data, processing of data, protection of the data, and various organizational processes. In addition, automation can potentially minimize obstacles of integrating multiple domains in an organization, while gracefully (i.e., without major disruptions to stakeholders) supporting changes to data and domains (e.g. addition or extension of new data products; modification of data or data products). For data analytics, Diagnostic, Descriptive, Predictive, and Prescriptive analytics solutions (Figure 3) are enabled to use the data to produce knowledge important to the stakeholders (e.g., measures; metrics; KPIs; etc.). This JDM incorporates a Registry API, Data Registry, Semantic Layer, and Data Layer components.

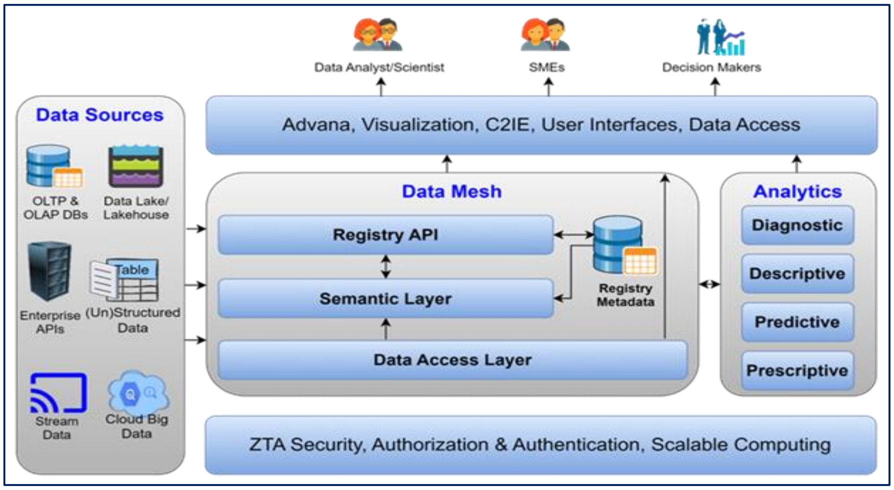

Figure 3. Proposed Federated Joint Data-Mesh Architecture

The Registry API component provides services required to register and manage data domains, products, and access, as well as other data functionality; It is a central data strategy element, enabling support of data analytics by providing easy access to data services. Automation of the registration process uses a standard interface to be called by the registry and implemented by each domain. The interface defines methods to collect the necessary information from the domain. Benefit highlights of the Registry API include:

- ❑ Provision to obtain required roles to access data products, as well as definition of the data products;
- ❑ definition of the data products, and functionality to help the registry manage and maintain domain metadata;
- ❑ the electronic means for the discovery of available data products published by the different domains;
- ❑ Searching capabilities allow the retrieval of data from different data products across domains.

The Data Registry attempts to simplify the management of centralized access to the organization's data by requiring the domain owners to drive most of the functionality. While there are similarities to a data catalog, a data registry does not centralize management of all the metadata and processes to realize the benefits. Also, the data registry facilitates both the provenance and ontology of data. Additionally, the registry defines a semantic layer, by which a consistent view of data products is provided; It allows access to its functionality through a user interface tool and/or API for programmatic access.

The Joint Data Mesh delivers a data-centric and diverse consumer-based oriented solution to best support data landscape changes, proliferation of data sources and consumers, stakeholder analytics needs, and agility to respond to change.

The Semantic Layer's unified and standardized ontology enables a consistent view and use of the data across multiple domains and does not preclude stakeholders from accessing the data products directly, nor add any significant overhead. Consistency is achieved using the normalized field name and values. Mapping the data product into the semantic layer is accomplished with the assistance of tools able to analyze the new product data to determine potential entities and fields they map in the ontology. As such, this architecture and ontology do not impose on the registration of new domains or data products to facilitate immediate accessibility for stakeholders.

The data mesh Data Layer is the workhorse of the architecture; It provides searching capabilities across all data products across different domains. Furthermore, this layer scales to stakeholders' demands and supports different types of data repository technology, such as SQL and NoSQL databases; data warehouses, data lakes, and data lake houses; structure and unstructured file repositories; streamed data repositories; cloud repositories; data APIs.

PROTOTYPE DISCUSSION

To explore the data mesh concepts this paper presents, we created a JDM prototype using the following JLVC Federation tools and systems: JTT, Joint Training Data Service (JTDS) Order of Battle Service (OBS), Low Overhead Training System (LOTS); Joint Exercise Control System (JECS). We developed a scenario using JTT for planning and design; A notional downed pilot scenario.

Joint System Utilization

For purposes of this unclassified prototype, we used JTT to develop training objectives based on universal joint tasks relevant to our scenario. Thus, our scenario is relevant to real-world operations. To note, as described in this paper's background section, for actual CCMD exercises, training objectives are created based on CCMD's Joint Mission Essential Task List.

JTT

Each UJT is coded with a unique identifier which is a key field for linking training objectives to other JTT data elements. In JTT, we developed an MSEL list and timeline (Figure 4)—a chronological timeline of pre-scripted events used by exercise controllers—to guide the exercise and training audience toward specific objectives. To note, MSELs can be categorized as simulation-supported or white-carded, and training objectives link MSELs to specific UJTs. For this paper, the downed pilot scenario, alongside a few preexisting unclassified JTT scenarios, was used to demonstrate the ability to query data across the JTT exercise repository, showing the linkage to a JLVC-supported event.

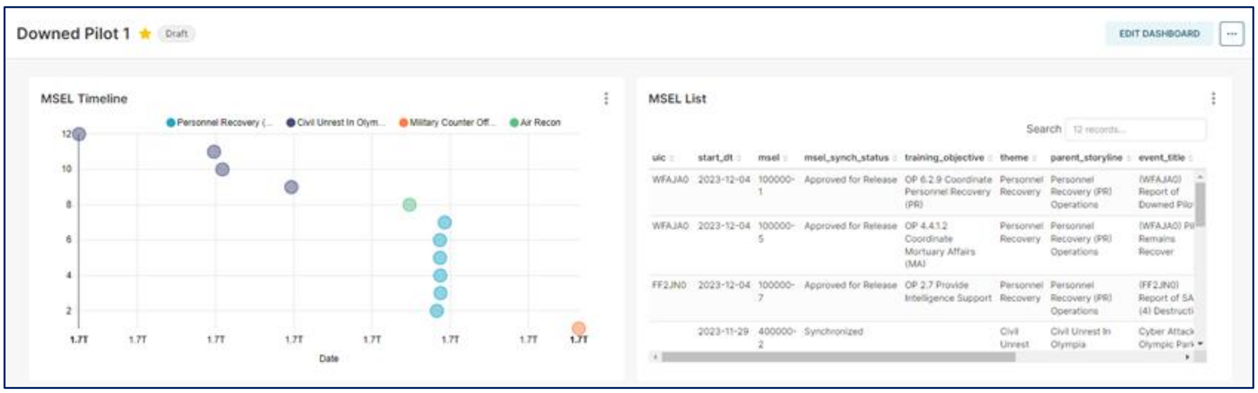

Figure 4. Notional Downed Pilot Scenario MSEL List and Timeline

JTDS OBS

We utilized the JTDS OBS to build out scenario forces for import into JTT. JTDS, a web-accessible application, can rapidly create a simulation scenario based on authoritative sources. There are two major services under JTDS:

1) The Terrain Generation Service (TGS) is built on a global terrain repository and geospatial applications;
2) The OBS is built on a repository of force structure data that includes units in an administrative structure, with associated entities (platforms, lifeforms, munitions, supplies, and emitters).

The OBS uses the UIC as the unit identifier and DIS enumerations for entities. Also, OBS provides a user interface to easily generate a scenario with sides, task-organized units based on scenario requirements, and map starting locations. Based on JTT MSELs, we developed a notional force structure in JTDS, with the UIC as the linkage to the MSELs.

JLVC LOTS Simulation Run

Many of the 34 JLVC Federation simulations and tools are developed by the Joint Staff and Service/Agency developers. For this prototype, we used LOTS to run the scenario using the JLVC-compliant obs.xml 4.3 schema. The scenario was run without monitoring by a simulation operator in real time and took approximately three hours. In an actual CCMD training event, a simulation operator would monitor the run to ensure results were realistic to the training audience. While LOTS was running the scenario, we used JECS to collect logs of the simulation actions and engagements. While this paper only focuses on the logs from a single small scenario, the principles would apply to multiple scenario logs of complex JLVC events involving many federated simulations, as they are all logged by the JECS application.

Demonstration Example

To demonstrate how existing Joint Training systems could be connected through a data mesh to provide metrics, we employed the Trino application to connect the event planning & design data in JTT, the force structure data in JTDS, and the simulation log files from JECS. Aircraft in this scenario include an original patrol with one UH-60M shot down, intel collection UAVs, F-16CJ Wild Weasels for surface-to-air-missile (SAM) site engagement, and UH-60Ms to transport Quick Reaction Force, with AH-64Es escorting. Figure 5 features aircraft tracks captured from scenario JECS log files and displayed in Trino. Additionally, Trino was able to capture metrics on the units used in the scenario, the types and numbers of munitions expended in the downed pilot vignette, and categorize the types of UJTs associated with the exercise.
Additionally, the prototype can process the log files to replay the portion of the simulation related to the cyber vignette.

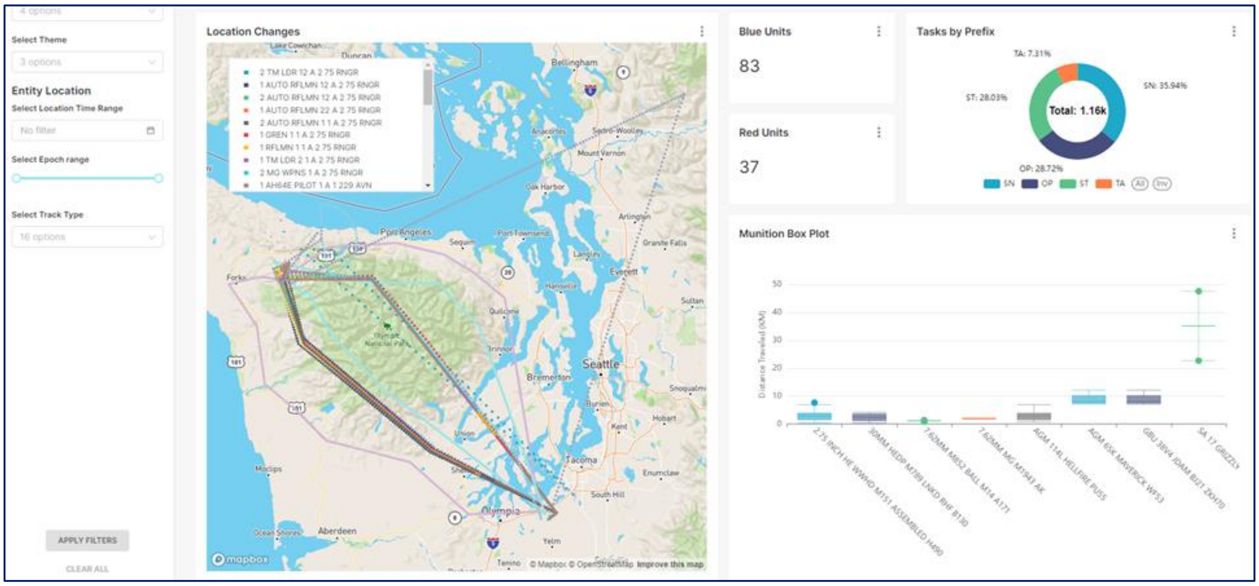

Figure 5. Aircraft Tracks Captured from Scenario JECS Log Files and Displayed in Trino

Trino facilitated easy data queries using SQL, as well as a display of key parameters on an easy-to-read dashboard. By leveraging key fields (e.g., UJT; UIC; DIS enumerations), it is possible to answer relevant questions such as, in the past five years:

- ❑ How many cyber-related MSELs were supported with simulation (vice white carded)?
- ❑ What training audiences were involved in the simulation-supported scenarios listed above?
- ❑ How many exercises were trained to cyber vignettes and what were the outcomes of the cyber simulation?
- ❑ Extract the associated observations and reports related to cyber vignettes from JTT.

WAY AHEAD RECOMMENDATIONS

Our demonstrated solution is founded on principles of adaptability, flexibility, sustainability, managing data at or nearest to the source, and unity of effort. This paper aligns with JLVC modernization efforts to address Joint Operational Training and Readiness Gaps, as well as NDS requirements to "implement institutional reforms that should integrate data, software, and artificial intelligence efforts to speed their delivery to the warfighter." Additionally, we acknowledge our recommendations hinge upon, writ large, the willingness across DoD to not fall into stove-pipes (E.g. single-platform with centralized data control methodologies). These risks are avoidable or mitigated by:

- Promoting a culture of willingness, early recognition, and collaborative steerage away from stovepipes;
- a Data Governance Framework promoting centralized oversight and decentralized execution – unity of effort;
- deliberately developing mesh-compatible architectures, with the following in mind: interoperability, documented data standards, data tagging, data sharing at speed and scale, open architectures, and vendor-lock avoidance.

Figure 6 illustrates a Data-Centric Organization's journey in terms of complexity/governance and technology/culture.

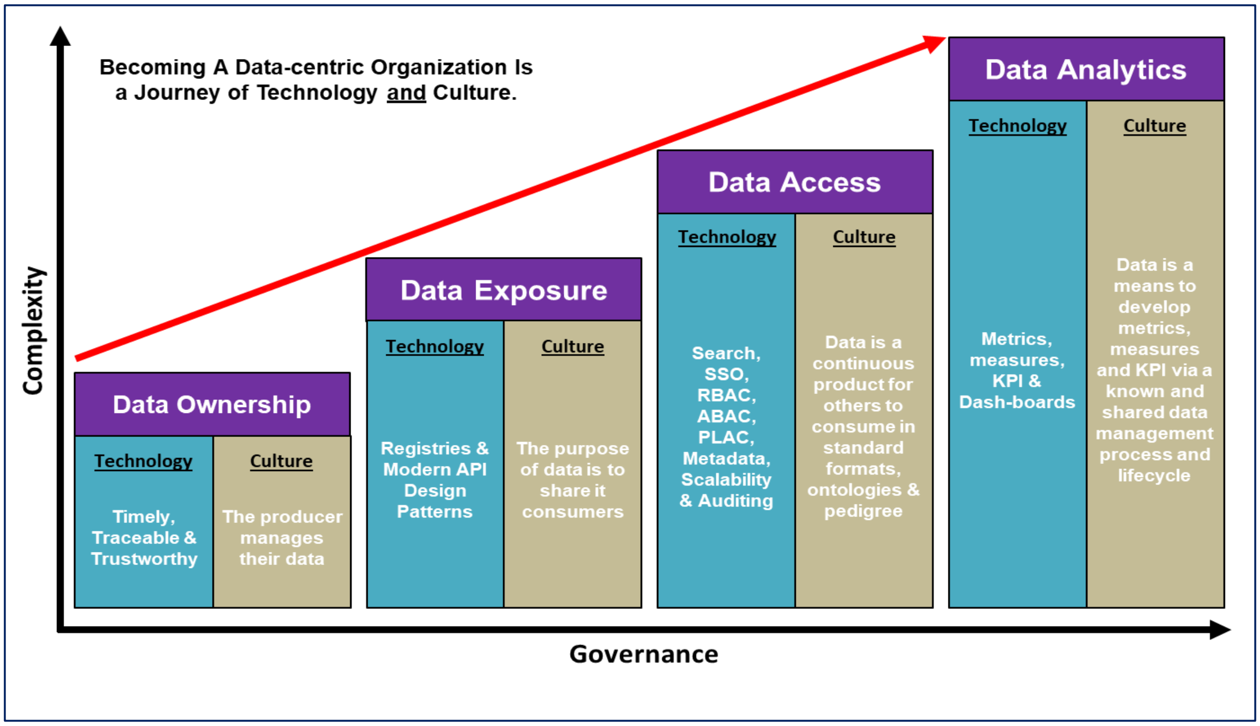

Figure 6. Data-Centric Organization's Path to Data Governance (Culture) and Architecture (Technology)

We recommend creating a Data Governance Framework to codify a measurable data strategy implementation plan and transitioning this paper's proposed Federated JDM Architecture from prototype to operational solution, to align with NDS priorities and DoD Data Strategy. These actions will enable a more comprehensive, data-driven assessment and reporting of force/unit readiness. A method to achieve this plan includes:

- Culture: Champion a centralized data governance framework—with data owner representation—and a decentralized data architecture, management, and execution at the point of origin/ownership. Institutionalize, integrate, leverage, and operationalize data—as a weapon system.
- Governance Program and Data-Fit-for-Purpose Framework: Promote quality, discoverable, and fit-for-purpose data, to enable data-driven measured-risk decision-making at the speed and scale of the fight.
- Technology and Decision-Support: Deliver value across one or more organizations and stakeholders, while reducing lifetime costs, minimizing data errors, and facilitating shared knowledge.

A data-centric JDM architecture for strategic readiness should include:

- A scalable, performant architecture—ZTA compliant with scalable computing; enables external to/from the data feed, MLS design (future support), and data security.
- Consumer Oriented—Self-service APIs for Senior Leader Decision Support, Joint All Domain Operations, and Business Analytics; web-based user interfaces with consistent views; metadata management.
- Discoverable, Measurable Data with Proactive Analytics—direct from the source and supported by unified, best-practice provenance and ontology; vendor-/tool-agnostic and minimized one-size-fits-all approach.

CONCLUSION

David Norquist stated, “Adversaries are also racing to amass data superiority, and whichever side can better leverage data will gain military advantage. Our ability to fight and win wars requires that we become world leaders in operationalizing and protecting our data.” This can be realized through a well-designed and executed joint data mesh and data federation that prioritizes digital transformation initiatives, data complexity, and bottleneck reductions, and on-the-fly adjustments over one-size-fits-all, single-solution (E.g. platforms; analytics; dashboards). Analogies:

- ❑ Joint force restructuring to only conventional forces with primary-specialized roles, divesting from highly trained special forces with multi-role unconventional expertise, or vice versa. Each has a purpose for which they are created and employed; The divestiture cost in lives, dollars, national security, etc. is contrary to calculated or measured risk decision-making.
- ❑ F-35 Joint Strike Fighter intended commonality benefits (E.g. spare parts; potential for combined maintenance training); however, there is a significant need to adapt to the carrier environment (E.g. carrier landings requirement for a tailhook, and reinforced landing gear and structural tolerance to sustain trapped landings).

As Abraham Maslow (1966) wrote, “If the only tool you have is a hammer, it is tempting to treat everything as if it were a nail.” While a single data architecture or data analytics solution may provide some benefits, the risks forced across the DoD, Allies, and partners could adversely hinder the data’s utility and strategic value. For example: a comptroller’s data, workflow, or tool requirements may not facilitate other organizations’/specialty areas’ unique needs for collection, evaluation, and visualization; It rapidly evolves to only facilitating the solution owner consumer prioritization at the expense of others. Moreover, single solutions create conditions favorable to vendors and adverse to consumers or decision-makers. Vendor-lock is especially detrimental if the solution is or becomes domain-/platform-centric vice agnostic; It allows significant dependencies, imposing excessive cost and technical mismatch issues when addressing sustainability and agility to pivot in line with dynamic technology growth and global security environments. Adverse 2nd/3rd order effects could include limited scalability and future modernization, unintentional data misuse or deletion, or MPE complications due to pre-committed mission partners' investment obligations.

A ‘data federation’ encourages a ‘Data-Centric Unity of Effort’; Multiple distributed data-centric architectures/databases functioning as one—synchronized, coordinated, integrated. In other words, the concept is harmonized efforts across multiple organizations and technologies advancing towards similar objectives.

Thereby, to best modernize JLVC—to become a modern Data-Centric Architecture and foster Analytics-Ready Organization transformations—we encourage a balance of organizational and domain autonomy, technology, and capabilities, supported by consumer-driven artificial intelligence (AI)/machine learning (ML) algorithms to:

- ❑ Be responsive to varied consumer group requirements (i.e. user-/decision-maker-centric);
- ❑ quickly adapt to rapidly changing technology and budgets;
- ❑ successfully share across the joint force and support a common JTSE;
- ❑ informs the organization, development, training, and readiness of Joint Forces.

Our approach aligns the JS J7 and JLVC federation with the NDS and DoDDS vision to transform into a data-centric organization that uses data at increased efficiency and at speed and scale for operational-, information-, and decision-advantage across the continuums of training, competition, and conflict. Moreover, our prototype and proposed approach, with its associated advantages addressed in this paper, delivers a capability that addresses Joint Operational Training Gaps and operationalizes mechanisms to provide consistency and coherency to JS J7 data decisions and enables CDS, MPE and ZTA, while improving data fitness, timeliness, and effectiveness. Finally, our solution scales to data landscape changes, new data producers (domains) and consumers (stakeholders), evolving analytics needs, and changes in the synthetic, informational, and operational environments.

REFERENCES

CJCSI 3500.01J, *Joint Training Policy for the Armed Forces of The United States of America*, 13 January 2020

CJCS I3401.01E, *Joint Combat Capability Assessment*, 13 April 2020, Current as of 19 May 2014

CJCSN 3500.01, *Chairman's Joint Training Guidance*, annual publication

CJCSI 3500.02C, *Universal Joint Task List Program*, 19 December 2022

CJCSM 3500.04F, *Universal Joint Task List Manual*, 01 June 2011 (note, at the time of this paper, version G is in staffing)

CJCSM 3500.03E, J*oint Training Manual for the Armed Forces of The United States of America*, 20 April 2015

CJCS Guide 3501, *The Joint Training System: A Guide For Senior Leaders*, 5 May 2015

Data & Analytics Center of Excellence Playbook, Government Service Administration IT Modernization Center of Excellence, October 2020, (https://coe.gsa.gov/docs/2020/Data%20Playbook-October2020.pdf)

Dehghani, Zhamak. "Data Mesh, Delivering Data-Driven Value at Scale." O'Reilly Media, Inc. March 2022

Description of the National Military Strategy, 2018

DoD Chief Data Officer, *DRAFT DoD Metadata Guidance Version 1.0*, July 2022

DoD Data, Analytics, and AI Strategy, Draft 2022

DoD Data Strategy, 30 September 2020

DoD Deputy Chief Information Officer's DoD Architecture Framework (DoDAF) Version 2.02. "DM2: DoDAF Formal Ontology. Accessed 04 February 2023. https://dodcio.defense.gov/Library/DoD-Architecture-Framework/dodaf20_ontology1/

Federal Data Strategy Framework, 2020. As of 02 November 2021, https://community.max.gov/display/DATA/Federal+Data+Strategy

Gibeling, K. E. (2022, May 12). *Universal Joint Task List (UJTL) 101*. [Powerpoint Slides]. Joint Staff, J7 Directorate.

National Defense Strategy of The United States of America, 2022

U.S. Government Federal Data Strategy website. "Federal Data Strategy Framework and Action Plans." Last accessed 04 February 2023. https://strategy.data.gov/overview/

U.S. Government Services Agency, IT Modernization Center of Excellence website. "Data & Analytics Center of Excellence Playbook." Last accessed 04 February 2023. https://coe.gsa.gov/docs/2020/Data%20Playbook-October2020.pdf