Data Mesh Reference Architecture
(DMRA)

Chief Digital and Artificial Intelligence Office (CDAO)

Version 2.6

March 2024

0

Contents

1. Executive Summary

The Department of Defense (DoD) Chief Digital and Artificial Intelligence (CDAO) Data Mesh Reference Architecture (DMRA) provides a blueprint to guide and constrain the instantiations of data mesh ("mesh") solution architectures. The DMRA is considered an organizational asset which:

- Provides common language for the various stakeholders
- Develops connection points among solutions' implementations
- Supports the validation of solutions against proven Reference Architectures (RA)
- Encourages adherence to common patterns

The DMRA is developed from an enterprise-level perspective in support of the Department's unified approach across data, analytics, infrastructure, and Artificial Intelligence (AI) activities. It uses a strategic approach to guide decentralized data management action across DoD, as outlined in the DoD Data, Analytics, and AI Adoption Strategy, to accelerate and scale decision advantage outcomes towards the Department's digital transformation goals.

As shown across the top of Figure 1, the DMRA supports two DoD Mission Areas: Boardroom and Battlefield. Actors in these Mission Areas, which may be human or machine, require capabilities such as those on the righthand side to the mesh. This need triggers mesh actors to develop capabilities by following some combination of action pathways such as those on the left. These actors and actions guide identification of Boardroom and Battlefield use cases for the DMRA.

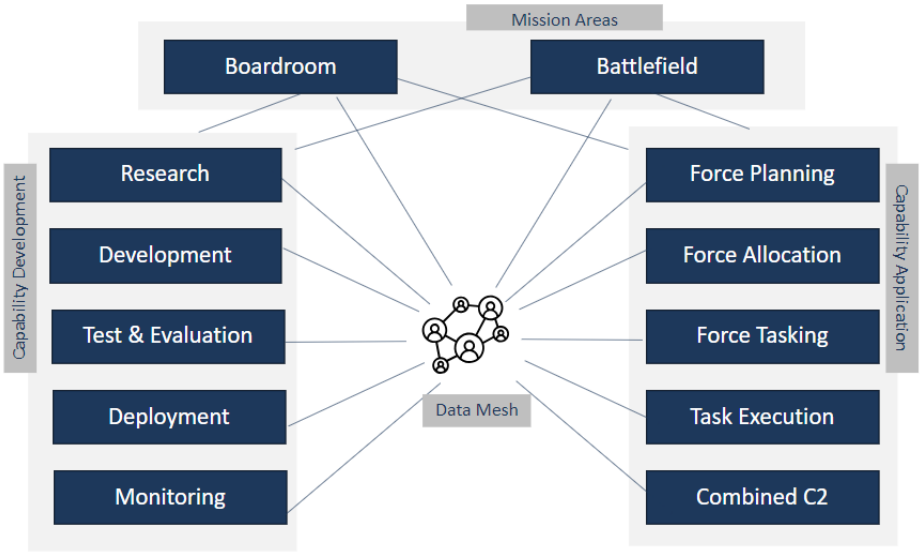


Figure 1 – DMRA Business Mission Areas

The DoD Architecture Framework (DoDAF) is the overarching, comprehensive framework and conceptual model enabling the development of architectures to facilitate the ability of Department of Defense (DoD) managers at all levels to make key decisions more effectively through organized information sharing[1]. Due to the scope of this document, only the relevant DoDAF models for the DMRA are presented, as shown in Figure 2.

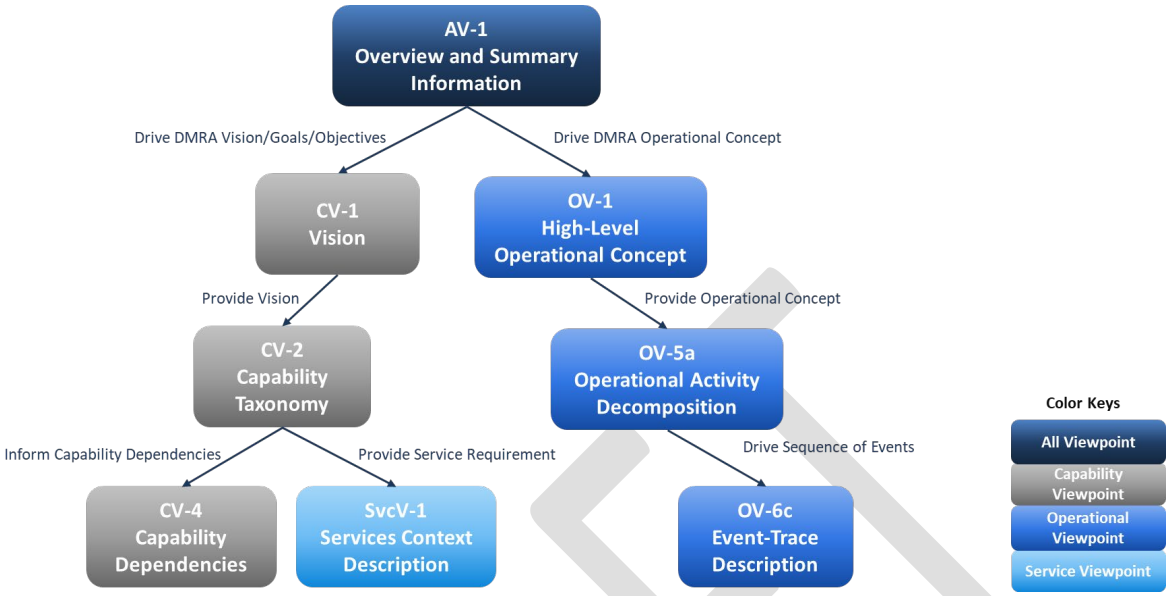


Figure 2 – Hierarchy of DMRA DoDAF Models

This document is organized to adhere the discipline of the DoDAF structure with the need to define use cases according to Figure 1. A mapping of the DoDAF models in the context of the use cases for Capability Development and Capability Application is illustrated in Figure 3.

Use Case / DoDAF Model	AV-1: Boardroom - Capability Development					AV-1: Battlefield - Capability Application				
	Research	Development	Testing & Evaluation	Deployment	Monitoring	Force Planning	Force Allocation	Force Tasking	Task Execution	Combined C2
CV-1: Vision	X				X	X				X
CV-2: Capability Taxonomy	X	X			X	X	X			X
CV-4: Capability Dependencies		X	X		X		X	X		X
OV-1: High-Level Operational Concept			X	X	X			X	X	X
OV-5a: Operational Activity Decomposition			X	X	X			X	X	X
OV-6c: Event-Trace Description				X	X				X	X
SvcV-1: Services Context Description				X	X				X	X

Figure 3 – Mapping of DoDAF Models in Document to Actors

2. Introduction

Purpose

The CDAO DMRA is the high-level framework to guide how Solutions Architectures and Systems Architectures (SAs) interoperate. It enables multiple strategy documents, including the DoD Data, Analytics, and Artificial Intelligence Adoption Strategy (2023), which states that "the Department will develop and implement a decentralized network among data providers and users. This network will

3

consist of both process-based and technical components, distributing ownership across data domains and treating data as a product."[2] That is a direct call for a mesh.

A mesh architecture is a decentralized approach that enables domain teams to perform cross-domain data analysis on their own[3]. It is widely known that no single mesh implementation will fit all organizations and that the mesh essentially makes existing SAs interoperable. As a result, DoD must design the operating elements, their functions, linkages, and how they fit within the DoD's mission context.

Appendix 1 All-Views (AV): AV-1 Overview and Summary Information displays an expanded executive-level summary of the DoD CDAO DMRA.

Strategic Approach

An RA differs from SAs in that the primary purpose of an RA is to guide and constrain the instantiations of the latter two[4]. The DoD definition for RA is:

> *Reference Architecture is an authoritative source of information about a specific subject area that guides and constrains the instantiations of multiple architectures and solutions.*

CDAO's strategic approach to developing the DMRA is to establish an understanding of both the status quo SAs and unimplemented approaches, propose mission context-sensitive best practices, and test their viability as components of the mesh in various scenarios.

The following actions are being taken to develop and prove the viability of the DoD CDAO DMRA, not necessarily in this order:

- Identify and develop relevant use cases that motivate the mesh and its design
- Determine the appropriate mesh design
    - Establish internal visibility: Gain a high-level understanding of SA instantiations to expose gaps and overlaps, thus informing practicality
    - Find external best practices: Probe best practices from both internal SAs and external partners
- Prototype where necessary: When a gap is identified for which no known best practice has a viable solution, prototype the solution internally with full documentation
- Experiment: As it is being constructed, iteratively test its practicality alongside mission partners

In summary, the strategic approach to the DMRA implementation is to take a global enterprise view and design what 'good' looks like so that individual SAs have something to strive towards. This document is rooted in months of evaluation of different SAs across many diverse partners and the resulting global and the local understanding of requirements, capabilities, and use cases.

The sections of DMRA were structured as follows to serve the overall purpose of this document:

- Section 1: Explained the DoD context for a data mesh.
- Section 2: Provided the purpose and strategic approach to developing the DMRA.
- Section 3: Provided overview of the use cases to motivate the development of the data mesh services.

- Section 4: Articulated the strategic purpose, principles, use case actors/roles, and architecture patterns for the DMRA to meet the CDAO strategic objectives.
- Section 5: Provided use case scenario walkthroughs to help understand how the user might interact with the data mesh by utilizing the mesh services.
- Section 6: Provided summary information of the DMRA and how the sections fit together and accomplish their goals to guide navigation through.

3. Overview of Use Cases

Use cases are written guidance for how a particular actor—human or machine—interacts with the data mesh to achieve a stated task. The following actors and use cases were carefully designed to illustrate the diversity of actors and the kinds of collaborations/interactions that can take place to achieve a goal in a complex and distributed data environment. This diversity encompasses data domains, data sources, formats, policies, and other factors, from the boardroom to the battlefield. The DMRA therefore develops the context for a relevant, feasible, and adaptable RA.

Boardroom Use Cases enact boardroom actors, who will ask data SMEs to query the mesh for empirical evidence to support decisions such as resource allocation and acquisitions. Examples include front office leaders such as the Deputy Secretary of Defense (DSD), and bots using large language models (LLMs) to propose courses of action (COAs). Please see the upper left of Figure 1.

Battlefield Use Cases enact battlefield actors, who will ask data SMEs to query the mesh for empirical evidence to support tactical decisions such as battle plan execution, logistic and sustainment, and machine-to-machine (M2M) edge decisions. Examples include frontline commanders, edge devices, and maintenance supervisors. Please see the upper right of Figure 1.

Actions refer to the capability development activities that are known to be performed by actors, include research, development, deployment, monitoring, and test and evaluation (RDT&E). Please see the lefthand column of Figure 1. These known actions are described below.

- Research: Anyone performing research will use the mesh in a highly exploratory manner to discover available data, learn how it is traditionally transformed, joined, and employed, and/or drill into specific questions. *Example: AI Engineer*
- Development: Anyone developing solutions will use the mesh to pose questions, develop dashboards and dictionaries. *Examples: UI/UX Professionals and Data Scientists*
- Test & Evaluation (T&E): Anyone doing T&E will use the mesh to verify, validate, and assess the value or appropriateness of "fit" solutions and decision outcomes. *Examples: Responsible AI Engineers in the Department of Test and Evaluation (DOT&E), and Data Scientists validating their results*
- Deployment: Anyone primarily concerned with supply chain operations. *Example: Tank Drivers relying on GPS for directions, tactical commanders for task execution*
- Monitoring: Anyone primarily concerned with ongoing sustainment and operations. *Example: Data Scientists*

Questions are the capability application queries that actors pose to the mesh. Please see the righthand column of Figure 1:

- **Force planning**: SECDEF (Secretary of Defense), OSD (Office of the SECDEF), CJCS (Chairman of the Joint Chief of Staff), JCS, JS (Joint Staff), and CCMDs (Combatant Commands) need data on the attributes and status of adversary, US, and coalition forces, and industry base capacity/capability to allocate resources to the development of new DoD platforms, weapons, systems, and software, as well as the readiness of its people, places, and things
- **Force allocation**: Since threats change rapidly in the year of execution, the SECDEF Book of Orders (SDOB) process leverages data on the status of adversary, US, and coalition forces, including capabilities, readiness levels, and planned modernization and training to inform how to respond to the latest threats
- **Force tasking**: In addition to the data fed to force allocation questions, CCMDs, Joint Task Forces (JTFs), and Service Component Commands (SCCs) assign tasks to the forces under their command and control require information on weapons load out, capabilities, and status, and logistics status to understand how to balance new and existing tasking in both connected and disconnected environments.
- **Task execution**: In addition to some combination of the above, JTFs, SCCs, and tactical commanders require fire control quality tracking and targeting data, and battle damage assessment data to match unit capabilities to mission needs, get the right approvals from legal and intelligence experts, and respond to rapid changes on the battlefield. These capabilities must be available in disconnected and degraded environments with a turnaround of anywhere from seconds (ideal) to days.
- **Combined Command and Control (C2)**: The United States (US) and coalition partners need to share data bidirectionally on status of adversary and forces, as well as data that can be used in their own targeting cycles. This question is called out separately because of the unique architecture needs of coalition collaboration and warfighting, but it must also be baked into the solutions for the use cases above.

A high-level notional walkthrough generalized to capture all of these use cases is in Section 5 of this paper. First, however, we will discuss the foundational 15 services that compose the DoD-wide data mesh.

Tactical Structure

Figure 4 lists the 15 following mesh capabilities that have been suggested as baseline to the mesh.

Service	Data Mesh Service Title	Service Description
1	UID	Tool to provide an enterprise wide globally exclusive and unique reference identifier to track any object
2	Semantic Services	Tools to promote sharing, collaboration and reuse of data model and ontologies; alias re-referencing to build a canonical controlled vocabulary.
3	Federated Data Catalog	Virtually federated catalog enabling Defense-wide visibility of data and interfaces through pointers to DoD assets and services.
4	Data and Metadata Profiles (xBOMs)	Managed service providing attribution (characteristics) that describe the meaning and intended use for data, metadata, algorithms, hardware, software and data objects (files, etc.).

5	Policy Access Control	Tools for ensuring proper access restrictions and identity verification for all consumers and producers in the data mesh.
6	Digital Policy Administration	Policy administration points feeding enforcement points enabling managed data access across environments.
7	Data Exchange Management	Handles and routes API requests to appropriate services.
8	Data Product Search	Tool for fast, relatable, and semantic understandable searching of all data products. Provides intuitive result finding for ingenuity and novel discovery of data products.
9	Data Mesh Pub/Sub	Systems of producers and consumers given by asynchronous service-to-service communication.
10	Mesh Performance Analytics	Track the flow and usage of data products across the mesh. Flow monitoring and alerting.
11	Data Product Life Cycle Management	Submits data products for registration to the domain and enterprise catalogs. Updates/maintains/revokes registration, as necessary. Manage recalled data products. Provide recall and other data product-associated notifications to data product consumers.
12	Data Security Classification	Microservice tools and policy for proper marking of all types of sensitive data across the DoD.
13	Quality Management Services	Tools for properly computing quality metrics on data as marking the data appropriately with its quality level.
14	Mediation Hub Services	Managed service for sharing and modifying producer data products into consumer's domain-driven context through translation services.
15	Mesh Instrumentation Tools	Behavior analytic data stream analytics to allow performance optimization and asset value determination.

Figure 4 Data Mesh Services

4. DoD-Wide Reference Architecture Definition

Strategic Purpose

The goal is to have a fully interoperable mesh that enables easily discoverable and accessible data across data domains to accelerate decision advantage at speed and scale [5]. The DMRA presents conceptual architectures of how to best orchestrate, or pull together, the DMRA elements. It does this by demonstrating that theoretical work performed to date is operationally viable versus a strongly-defined structure.

The DoD enterprise benefits from the DMRA in multiple ways. First, it achieves clarity beyond the high-level strategic work that is being performed by others on what the mesh will look like. This gives each CDAO directorate the tools to know how they will interact with it. Second, it is rooted in tactical knowledge of best practices in the latest offerings, both internal and external in industry. This awareness is critical to optimizing the mesh design for DoD's unique mission set. Third, where necessary, the DMRA guides the design of SAs.

There is no single 'correct' mesh implementation because each organization and mission has different requirements, idiosyncrasies, and resources. Thus, the real challenge is to develop a responsible and decomposable framework for how components together versus developing strongly defined standards.

The DMRA is positioned to align with the following CDAO strategic objectives [2] [3]:

- Improve Foundational Data Management *through Domain Ownership & Data as a Product*: Increase the quality and availability of relevant DoD data to support advanced analytics and AI capabilities for both business and warfighting functions to make rapid, well-informed decisions
- Deliver Capabilities for Enterprise and Joint Warfighting Impact *through Domain-Oriented Decentralization for Analytical data*: Enhance and/or generate warfighting capabilities with data, analytics, and AI technologies for improved decision advantage outcomes
- Strengthen Governance and Remove Policy Barriers *though Federated Computational Governance*: Ensure responsible behavior, processes, and outcomes while accelerating the pace of adoption for data, analytics, and AI technologies across the Department
- Invest in Interoperable, Federated Infrastructure *through Self-Serve Data Infrastructure Platform*: Optimize the Department's federated infrastructure to support scaling data, analytics, and AI adoption and improve interoperability
- Advance the Data, Analytics, and AI Ecosystem *through Data Mesh Architecture*: Strengthen intergovernmental, academic, industry, and international partnerships to enable adoption of data, analytics, and AI technology
- Expand Digital Talent Management *through Enhanced Technical Foundation*: Increase hiring, training, and retention for the most critical data, analytics, and AI-related work roles

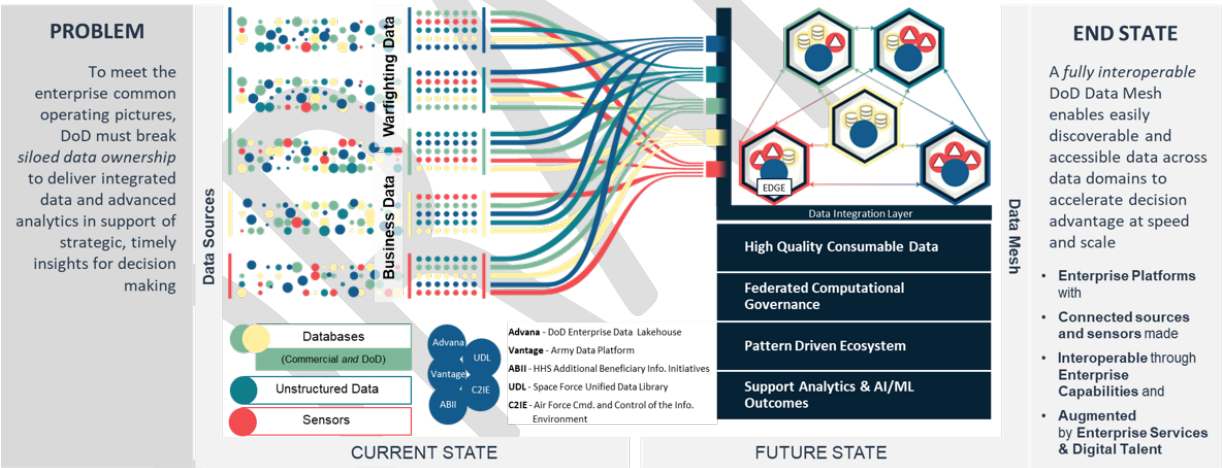


Figure 5 – CV-1 Vision

The Capability Viewpoint (CV) CV-1 Vision in Figure 5 provides the strategic context for the capabilities described in the DMRA. It communicates the strategic vision for the capability areas fulfilled by the mesh and describes how the strategic vision and high-level goals and objectives should be delivered to overcome the strategic challenge on the lack of data interoperability as identified in the problem statement.

Starting from the far left, the initial problem statement is that data is siloed and its interoperability impaired by inability to fluidly translate standards and languages. The colored dots in the center indicate that data may come in many different sources, formats, shapes, and sizes. They may be institutional databases of structured or unstructured data, stream directly from sensors and edge devices, or be standard data used for model training. Regardless of source and format, the central pipelines make them

8

accessible to any appropriate endpoint, such as CDAO's Advana or Army's Vantage. Each of those employs access and privacy controls to grant its users access to the data for that which is their specified end goal, e.g., reporting dashboards, edge-to-edge device interoperability, and AI training.

The Operational Viewpoint (OV) OV-1 High-Level Operational Concept Graphic in Figure 6 depicts how data and data teams utilize the operational concepts and unique aspects of the DMRA. Starting from the left, the various mission-specific, heterogenous data platforms indicated by the four parallel horizontal pipelines of colored dots in the CV-1 are transformed for specific questions, or Data Domains. Mission-specific Data Product Teams (DPTs) aggregate those data and data assets into Data Products (DP), defined as logical units that contain all components to process and store domain data for analytical or data-intensive use cases and makes them available to other teams.

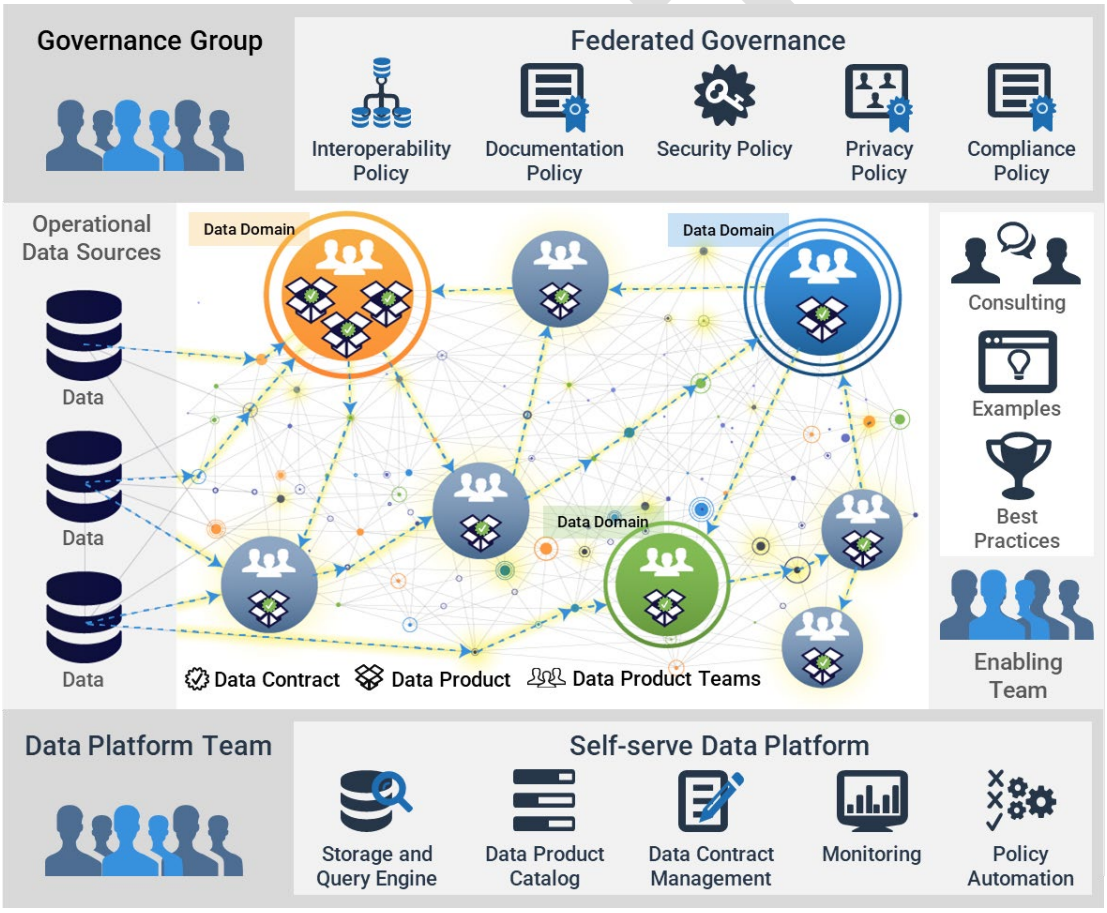


Figure 6 – OV-1 High-Level Operational Concept Graphic

Across the bottom, the DPTs validate data quality and manage Data-as-a-Product thinking with Data Contracts and Policy Automation, thus making the DPs available to self-service users who now trust that the data has been treated appropriately for their use case.

The top horizontal indicates the multiple governance and access functions monitoring privileges. The Governance Group defines DP Quality via Federated Governance by dictating interoperability,

9

documentation, security, privacy, classification, and compliance policies. Thus, when a DP complies with the policies agreed upon by the governance group it is readily consumable. The DPTs define and maintain the data quality and core data mesh services bespoke to each mission use case. The core services allow the DPTs to administer data products and issue policy by the governance group while enabling users to subscribe to data products.

On the far right, the Enabling Team is a federated group of data platform representatives that support domain teams and consumers. They are the data mesh advocates that support the adoption of the data mesh across the enterprise.

Importantly, the actors performing the actions illustrated in these diagrams may be human or machine. This is a critical element to keep in mind throughout the DMRA, wherein the framework facilitates machine-to-machine (M2M), human-to-machine (H2M), and human-to-human (H2H), thus enabling DoD's need for Coalition Joint All Domain Command and Control (C-JADC2).

Principles

Generic Mesh Principles

Alluded to the Strategic Purpose section above, the following four foundational principles of a mesh are described in detail below[3]:

- Domain ownership mandates the domain teams to take responsibility for their data, analytical data should be composed around domains, and analytical and operational data ownership is moved to the domain teams, away from the central data team
- Data as a product projects a product thinking philosophy onto analytical data, i.e., there are consumers for the data beyond the domain. The domain team is responsible for satisfying the needs of other domains by providing high-quality data. Basically, domain data should be treated as any other public API
- Self-serve data infrastructure platform adopts platform thinking to data infrastructure. A dedicated data platform team provides domain-agnostic functionality, tools, and systems to build, execute, and maintain interoperable data products for all domains. With its platform, the data platform team enables domain teams to seamlessly consume and create data products
- Federated computational governance achieves interoperability of all data products through standardization, which is promoted through the whole data mesh by the governance group. The main goal of federated governance is to create a data ecosystem with adherence to the organizational rules and industry regulations, allowing local variations responsive to particular needs.

A fifth foundational mesh principle that we added, mesh performance monitoring, utilizes tools for properly computing quality metrics on the performance of the mesh by conducting behavior analytic data stream analytics to allow performance optimization and asset value determination, allowing the mesh to self-identify aberrations and, when possible, to be self-healing.

Figure 7 identifies how the 15 data mesh services align to the aforementioned mesh principles.

Service No.	Data Mesh Service	Domain Ownership	Data as a Product	Self-Serve Data Infrastructure Platform	Federated Computational Governance	Mesh Performance Monitoring
1	UID	X	X			
2	Semantic Services			X		
3	Data Mesh Catalog			X		
4	Data and Metadata Profiles (xBOMs)	X	X	X	X	
5	Policy Access Control			X	X	
6	Digital Policy Administration			X	X	
7	Data Exchange Mgmt			X		
8	Data Product Search			X		
9	Data Mesh Pub/Sub			X		X
10	Mesh Performance Analytics			X		X
11	Data Product Life Cycle Management			X		X
12	Data Security Classification				X	X
13	Quality Management Services			X		X
14	Mediation Hub Services			X		X
15	Mesh Instrumentation Tools			X		X

11

Figure 7 – Data Mesh Services to Data Mesh Principles Matrix

DMRA Use Cases and Actors/Roles

Use cases develop the context for the mesh for different actors and actions to fulfill the functional area missions. Actors in the Boardroom and Battlefield have capability applications such as those on the righthand side of Figure 1. This triggers mesh actors, human or machine, to develop a response by following some combination of action pathways such as the capability requirements on the lefthand side. These actors and actions guide identification of boardroom and battlefield use cases for the DMRA.

Human input and output create the value of a data mesh. Therefore, it is essential to identify the actors and their roles in support of the use cases since the mesh is a decentralized sociotechnical approach to share, access, and manage analytical data in complex and large-scale environments within or across organizations[1].

Defining a system's actors and their roles is important for developing use cases of a system because it identifies the main applications, requirements, functionalities, and interactions of the system. Actors (human or machine) are the external entities that interact with the system. Roles are the specific responsibilities or tasks that actors perform in relation to the system. By defining actors and roles, we can specify what each actor expects from the system and what the system expects from each actor. This way, we can design use cases that describe how the system meets the needs and goals of the actors.

- Domain Team[6]: a cross functional team of data professionals focused on developing data discoverable and easily-utilized products.
    - Data Product Owner: oversee the development and delivery of data products. Defines the vision, scope, and roadmap of the product, as well as prioritizing and managing the backlog of user stories and tasks. Collaborates with Data Engineers, Data Scientists, analysts, and other cross-functional teams to ensure the quality, usability, and alignment of the product with the enterprise goals and needs
    - Data Engineer: creates and maintains DPs that serve a specific domain. Ensures that the DPs products are reliable, secure, scalable, and accessible to the data consumers. Collaborates with other data engineers and data product owners across the data mesh to share best practices and align on common standards
- Data Consumer [6]: performs analysis, makes decisions, or creates products or services. Primary role is to access and explore data that is relevant, reliable, and understandable for their needs. May use various tools and methods to query, visualize, or manipulate data, depending on their skills and goals. This actor can come in many forms for various use cases. i.e. Service or Robot Accounts, ML Engineers, Data Engineers, Data Scientists or Analysts
- Data Platform Engineer [2]: builds and maintains the data infrastructure and platform that enables the data mesh. Provides tools and frameworks for data governance, quality, and interoperability across the data mesh
- Data Governance Group [2]: defines the standards, policies and best practices for data quality, security, privacy, and interoperability across the data mesh. Monitors the performance and compliance of the data domains
- Enabling Team [2]: spreads the idea of data mesh within the enterprise. In the beginning of data mesh adoption, a lot of explanatory efforts will be required, and the enabling team can act as

data mesh advocates. Help domain teams on their journey to become a full member of the data mesh. The enabling team consists of specialists with extensive knowledge on data analytics, data engineering, and the self-serve data platform.

Capability Patterns

The CV-2 Capability Taxonomy in Figure 8 creates a hierarchy of capabilities and taxonomies by codifying the capability elements. It facilitates development of implementation details and capability audits and gap analyses by decomposing each 'parent' capability into its 'children' capabilities.

The following parent capabilities correspond to the 15 proposed capabilities in Figure 4.

1. Unique Identifier (UID): Function is to generate a unique sequenced UID inside the applicable domain, e.g., US Army (the domain) could have three sequenced UIDs for Active Duty, Reserves, and State National Guard. The capability dependencies include 1) systematic notification that a new UID is required, 2) ability to generate a new child UID inside of that domain for the three children listed above, and 3) trigger to generate a unique sequential child ID recognizable by multiple domains
    a. Vignette: A user must validate that a digital asset was used to generate a specific analytical result. The need for certainty requires that the identified asset is incontrovertible and will guarantee the same results are derived using the same process and algorithms
    b. Capability: The UID service is used to generate unique identifiers that are non-collided and persistent across time and ascribed to any object
    c. Outcome: The UID ensures that each instance of any object can be discretely managed and identified with a high level of confidence
    d. Acceptance Criteria: Each UID is unique, persistent across time, non-reusable, machine-readable, machine-comprehensible, and provides a reference index that is recorded and managed for asset inventory control
    e. Primary Interdependencies: Practically all other services rely on UID as a core service that makes different systems interoperable, however, UID is most tightly coupled with Semantic Services and [x]BOMs while others have secondary and tertiary dependencies
2. Semantic Services: Upon receiving an element of any given language set, identify whether that element exists elsewhere in the enterprise's known vocabularies. Notably, this includes comparing to federated dictionaries, not just a central standard dictionary. Multiple steps go into making this possible: the service must first receive the contextualized term and definition, then compare against the existing lexicons. It will either seed a new term or trigger review by the CCV governance body. Next, it will publish to the catalog and scan for additional collisions.
    a. Vignette: A Contracting Officer must be certain that the information being presented in a contract to an external party is precisely understood for both meaning (definition) and intent (context)
    b. Capability: The Semantic Services will manage each controlled vocabulary to ensure a complete understanding of each term is conveyed with precise meaning within a given context by presenting a moderately atomic definition that is scientifically defensible, unique, non-collided, and non-deviated

13

c. Outcome: A single *canonical* controlled vocabulary (CCV) is managed at the enterprise level, and multiple *local* controlled vocabularies (LCV) are managed throughout and beyond the enterprise. CCV and LCVs are independent, allowing for language-agnostic mapping based on definitional congruence and contextual alignment
d. Acceptance Criteria: The CCV is singularly unique across the enterprise. Each LCV is constrained to the local context and domain. A term, regardless of CCV or LCV, shall be:
    i. non-collided and non-deviated
    ii. unique regardless of language structure
    iii. assigned a machine-generated UID (see UID service)
    iv. assigned a machine-generated term (same model as the UID, but unique numerical space)
    v. assigned a scientifically defensible and moderately atomic definition associated to one or more contexts, where each context has a context description that explains why the definition is mapped to the context
e. Primary Interdependencies: Particularly important to Data Product Search, the Federated Data Catalog, and Mesh Performance Analytics. It is tightly coupled with UID and [x]BOMs

3. Federated Data Catalog: This is perhaps the best-understood of all the services as there are existing services available in DoD. The data mesh catalog is a reference containing metadata about DoD's data. When well-constructed and combined with CCV, UID, and BOMs, it enables searching across all referenced data assets (including data products). The definition of 'well-constructed' indicates an organized, governance compliant, policy aligned listing of all data assets, recorded in the catalog with all necessary characterizations for discovery.
    a. Vignette: A researcher is looking for all iceberg data that has been assembled for the past 20 years.
    b. Capability: The Federated Data Catalog can be queried to identify all the objects that have any content relevant to iceberg data. The results can be presented to a UI/UX for local inspection or presented to an information exchange mechanism to transit the findings to another capability.
    c. Outcome: The catalog stores data references that enable a consumer to target information based on characterizing query and then presents the findings to an appropriate presentation mechanism.
    d. Acceptance Criteria: The catalog capable of storing reference data about objects that can be discovered by direct or indirect search techniques. The search techniques shall include formal query methods that use characterizing metadata for locating objects. The exact contents of the catalog are dependent on the design and are evaluated based on the successful discovery and location of the object being sought.
    e. Primary Interdependencies: Due to its federated state, is tightly coupled with UID, CCV, and [x]BOMs to enable easy searching via the Data Product Search. The Federated Data Catalog feeds ingests Policy Access Control and feeds Mesh Performance.
4. Data and Metadata Profiles: A machine-readable (e.g., JSON, XML) listing of all the parts, subcomponents, and assemblies making up an asset. Enables dynamic management of assembly and sub-assembly, asset discovery, and insight into lineage, provenance, and pedigree. Analytic techniques to be used on [x]BOMs range the gamut from a simple search to advanced AI

14

a. Vignette: An analyst requires an exact definition and full characterization of a platform to explore recurring but isolated performance deviations.
b. Capability: The Data and Metadata Profiles service retrieve and expose the platform's componentry, including part version and origin, in a machine-readable and machine-comprehensible presentation format. The service must manage all three types of profiles which include:
    i. Atomic object characterization - A Bill of Materials (BOM) that stores all necessary metadata to fully describe the object. *(Type 1 BOM)*
    ii. Composite object characterization – A BOM that stores all necessary metadata to fully describe each discrete atomic object and any other composite object characterization BOM (BOM of BOMs) necessary to describe the complete composite list of all items in the composite. *(Type 2 BOM)*
    iii. eXtensible process model characterization – A BOM that includes Type 1 and Type 2 BOMs and one or more process model descriptions that describe sequencing events that process an outcome.
c. Outcome: The full characterization of atomic objects, composites, and processes in machine-readable, machine-comprehensible expressions easily used by a consumer of a data object or processing sequence without any further assistance. The analyst is now able to perform outlier analysis on the specific platforms with performance deviations.
d. Acceptance Criteria: The service can generate and manage BOMs that are pattern and standard compliant and enable well-formed characterizations. The characterizations can be used by a consumer to provide a full understanding of data objects or reconstitute a processing sequence.
e. Primary Interdependencies: Tightly coupled with UID and CCV to enable easy translation from one system to another, for instance, when one system identifies parts via Serial Numbers (SERNO) and other identifies with part numbers. dataBOMs and aiBOMs enable transparency into the mesh via Mesh Performance Analytics and also monitor access and privacy via Policy Access Control and Digital Policy Administration.

5. Policy Access Control: Manage access to data products using metadata about entity actors paired with digital policy administration (see later).
    a. Vignette: A data producer has restricted data product to a specific community.
    b. Capability: The Policy Access Control service enforces the access control disposition created by using computational rule logic driven by the specified attributes.
    c. Outcome: Mesh objects (data and resources) are accessible if and only if the attributes and rule logic compute a favorable access control disposition.
    d. Acceptance Criteria: The rule logic and applied attributes result in a repeatable and predictable binary disposition of grant or deny access. The enforcement of access control is irrespective of the environment.
    e. Primary Interdependencies: Policy controls are tightly coupled with Digital Policy Administration and are a dependency of Data Product Search, Federated Data Catalog, and [x]BOMs.
6. Digital Policy Administration: The Digital Policy Administration service manages the creation, maintenance, and auditing of digital policies.

15

a. Vignette: A new data policy is issued by the DoD, and a digital enforcement policy must be produced for enterprise implementation.
b. Capability: The Digital Policy Administration service provides a logic rule authoring capability, a test and evaluation environment to simulate outcomes, and a publishing service to load the Policy Access Control service with the new digital policy.
c. Outcome: New rules or existing rule modifications have a UI/UX and authoring environment to create and test digital policy that is implemented based on written policy.
d. Acceptance Criteria: The rule logic can be created using primitives or open rule expression symbolics to create logic patterns that use attributes to derive a behavior that is enforceable using the Policy Access Control service. The rules and attributes produce a repeatable and predictable outcome based on all things being equal.
e. Primary Interdependencies: Works closely with Policy Access Control and [x]BOMs. For the latter, the logic rules dictated in the Digital Policy Administration are encoded in a machine-readable format that is easily consumed by machines to autonomously manage the initial pass at how to implement policies.

7. Data Exchange Management: considers all data exchange limitations known by DoD, from widely-accessible (Application Programming Interface [API]) to cloud storage solutions to access-denied environments such as ships at sea or satellites in space. A programmatic optimization of which method is best for any given use case.
    a. Vignette: A data object is registered into the Federated Catalog. This new data object has a unique structure and content composition.
    b. Capability: The Data Exchange Management service manages the automated extraction, transport, delivery, and re-assembly from an origin repository to a consumer environment.
    c. Outcome: A consumer can invoke an Exchange Mechanism associated with a data object in a repository and have a data object delivered for use.
    d. Acceptance Criteria: The Data Exchange Management service is capable of registering multiple exchange mechanisms, publishing the inventory and all associated data objects, and execute deliver requested objects. The service must be able to present the pattern and standards for successful connection/use of each exchange mechanism and data object pairing, any constraints, and levels of performance and service guarantees.
    e. Primary Interdependencies: Relies on highly-functioning Digital Policy Administration, Data Security Classification, and Policy Access Control. Additionally, interacts with Data Product Search and the Federated Data Catalog to present required products to users.
8. Data Product Search: Critically, this function includes any data products along the continuum from raw, untouched data products to dashboards to advanced AI algorithms. The search function is critical for usability. It enables entities to find existing products, using anything from deterministic to stochastic searches on metadata and content. Once access is positive, all returned data is available in a view for users to work with.
    a. Vignette: An AI engineer is creating a new model to address a specific scenario, and there is a need to identify all relevant data across the enterprise.
    b. Capability: The Data Product Search service provides the ability to "scan" any and all data repositories. The search capability includes non-formal repositories such as SharePoint disk caches, mounted local share drives, and other sources of data.

16

c. **Outcome:** A consumer can obtain data from any "connected source". This does not mean the source is useful or well-described.
d. **Acceptance Criteria:** The ability to find data from any connected source is possible. This is limited by access control considerations.
e. **Primary Interdependencies:** Interacts with CCV, UID, [x]BOMs, Digital Policy Administration, Federated Data Catalog and Policy Access Control.

9. **Pub/Sub:** Enables movement of information among entities without them being aware of each-others' existence, thus overcoming the challenge of establishing individual point-to-point connections. It is triggered when parametric conditions are satisfied and happens asynchronously, meaning the communication does not rely on a specific channel to be open because it incorporates an intermediary, the pub/sub message broker.
    a. **Vignette**: A data producer repeatedly releases updates, accompanied by notifications to a data set and consumers.
    b. **Capability:** The commercially established publication/subscription capabilities are available in the mesh.
    c. **Outcome:** A consumer can subscribe to any published data object in the mesh.
    d. **Acceptance Criteria:** Each published data object in the mesh is registered to a Pub/Sub service and consumers may subscribe commensurate with access control restrictions.
    e. **Primary Interdependencies:** Consumers must be cleared by Policy Access Control, Digital Policy Administration. Furthermore, the data products themselves must be searchable in the Data Product Search and Federated Data Catalog with the Mediation Hub Services brokering translation from the original data product to the end consumer.
10. **Mesh Performance Analytics**: Monitors mesh usage to report on usefulness, gaps, and redundancies in its structure to identify areas for improvement.
    a. **Vignette**: Congress requires DSD to report on the value proposition of the Data Mesh investment.
    b. **Capability:** The Mesh Performance Analytics services provide programmable telemetry attributes for data collection based on each mesh service, data population, data requests, delivery volumes, request frequency, etc.
    c. **Outcome:** Data mesh behavioral producer and consumer analytics can be generated based on mesh activity.
    d. **Acceptance Criteria:** Analytic reporting is possible on any instrument-able focus
    e. **Primary Interdependencies:** Data Product Lifecycle Management feeds directly into the performance analytics via Data Exchange Management. The Mesh Performance Analytics is populated into a data product, which populates into the Data Product Search and Data.
11. **Data Product Lifecycle Management**: A key component to the mesh, creates a full-circle view of how data is being used, publishes end products back to the data catalog to be searchable by other users, and establishes accountability for users, both human and machine. Lineage management is a specific feature of this capability.
    a. **Vignette**: An auditor is investigating when a data error occurred and is required to do a time-phased impact analysis.
    b. **Capability:** The Data Product Lifecycle Management service tracks every data object and resource of the mesh from the point of origin or instantiation to retirement or removal, and every change that occurs over the entire lifetime.

17

c. **Outcome:** A complete history of each data object and mesh resource is tracked and available for inspection.
d. **Acceptance Criteria:** All changes, including instantiation and deletion of any data object or mesh resources, is tracked, recorded, discoverable, and reportable. The policy decision of what level of management and retention has no impact on the required capability.
e. **Primary Interdependencies:** Relies on UID, CCV, and [x]BOM services. Validates privileges and usage against Policy Access Control, Digital Policy Administration, and Data Security Classification.

12. **Data Security Classification:** Renders classification markings as a data element inside the data catalog and [x]BOM. This service proposes classification escalation as data is aggregated and algorithmic logic is applied, thus giving the otherwise-bland data intelligence potential.
    a. **Vignette**: An Original Classification Authority designates a data object as Top Secret (TS).
    b. **Capability:** The Data Security Classification (OCA) service generates all of the necessary ”marking” attributes based on the Classification and Material Handling designations, or other constraint models that are required. The attributes and the specific attribute values are injected into the appropriate profile(s) for the data object.
    c. **Outcome:** A data producer, data steward, or data product manager can specify all of the necessary control and constraint markings, and the service will generate the attributes and values, and then inject them into all of the appropriate profiles.
    d. **Acceptance Criteria:** The capability can manage all control and constraint markings and all allowed values and inject them into one or more profiles or generate a specific control profile that is associated.
    e. **Primary Interdependencies:** Tightly coupled with the CCV and [x]BOM capabilities. Interacts with the Digital Policy Administration and Policy Access Control capabilities.
13. **Quality Management Services (QMS):** Supporting lineage and provenance, this service ensures that data is appropriately marked and tagged for users to discern its applicability.
    a. **Vignette**: A medical study of toxicity in a region is being used to determine the course of treatment for ailing personnel, requiring accuracy and privacy considerations (HIPAA [Health Insurance Portability and Accountability Act], PII [Personal Identifiable Information], and PHI [Protected Health Information]).
    b. **Capability:** Provides the ability to “grade” data based on specific precision and accuracy considerations. The information is captured as attributes and recorded as metadata.
    c. **Outcome:** A consumer can review the quality scores of the data and determine if the data provides sufficient fidelity, precision, and accuracy are suitable to meet the analytical or compositional needs required to achieve the necessary outcomes.
    d. **Acceptance Criteria:** Captures attributes that characterize the features of the data object providing quality scoring. The scoring attributes are then injected into the appropriate profiles to record the scoring outcomes.
    e. **Primary Interdependencies:** Mediation Hub Services, Digital Policy Administration, Data Security Classification.
14. **Mediation Hub Services**: The federated nature of the mesh allows different entities to declare and use their desired data standard, however, when data is utilized in the mesh it must be transformed into a common pattern understandable by all, then, upon delivery to another user,

18

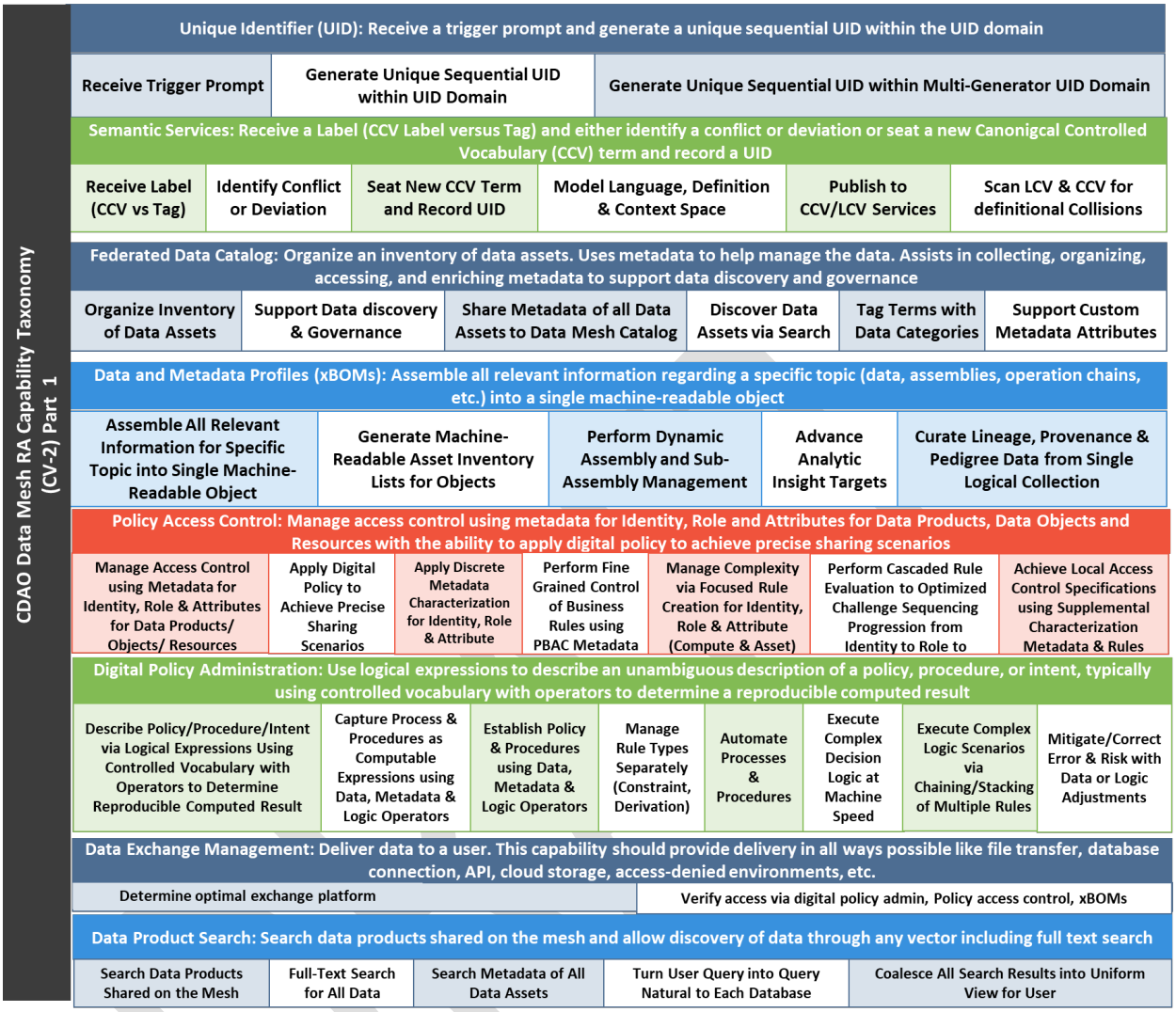
be reformatted into that user's desired standard. The interoperability patterns improve speed and transparency.

a. Vignette: A data object with high value content is stored in a format and language that is unique and unknown to the consumer.
b. Capability: The Mediation Hub Service provides the ability to prescribe the consumer target structure, formatting, and language or expressional notation and then transform the data object to be directly usable by the requesting consumer.
c. Outcome: A consumer can use a data object that is stored in a way that is not viable for use with the current consumer's capabilities.
d. Acceptance Criteria: The Mediation Hub Services can take a data object's structure and contents and translate and/or restructure the contents into a prescribed form. The content and values of the fundamental data object must retain the established quality score and not degrade the informational value in any dimension.
e. Primary Interdependencies: [x]BOMs and CCV to convert into local vocabulary and pattern. Digital Policy Administration to comprehend different policy patterns. Pub/Sub for receiving and issuing messages. Data Security Classification to convert classification definitions among partners.

15. Mesh Instrumentation Tools: Digital bots and other monitoring mechanisms for ensuring maximum up-time of the mesh, identifying, and isolating vulnerabilities, executing upgrades and rollbacks of software and hardware, and synchronizing with partners' toolkits.
    a. Vignette: The CIO is interested in understanding if there are opportunities to improve mesh componentry.
    b. Capability: The Mesh Instrumentation Tools provide IT-level telemetry data that can be used to assess infrastructure health and operational performance analysis.
    c. Outcome: The information can be collected to analyze the performance profile of discrete mesh elements.
    d. Acceptance Criteria: Each mesh component instruments in a way that allows management and executives to make data driven decisions on where to invest in mesh improvements.
    e. Primary Interdependencies: Payload on Mediation Hub Services and Data Product Search informs asymmetric workloads. Data Exchange Management and Data Product Lifecycle Management track data usage across its lifespan to watch workload drift across time.

See Figure 8 for an illustration of the above narrative.

19


CDAO Data Mesh RA Capability Taxonomy (CV-2) Part 1
Unique Identifier (UID): Receive a trigger prompt and generate a unique sequential UID within the UID domain
Receive Trigger Prompt
Generate Unique Sequential UID within UID Domain
Generate Unique Sequential UID within Multi-Generator UID Domain
Semantic Services: Receive a Label (CCV Label versus Tag) and either identify a conflict or deviation or seat a new Canonigcal Controlled Vocabulary (CCV) term and record a UID
Receive Label (CCV vs Tag)
Identify Conflict or Deviation
Seat New CCV Term and Record UID
Model Language, Definition & Context Space
Publish to CCV/LCV Services
Scan LCV & CCV for definitional Collisions
Federated Data Catalog: Organize an inventory of data assets. Uses metadata to help manage the data. Assists in collecting, organizing, accessing, and enriching metadata to support data discovery and governance
Organize Inventory of Data Assets
Support Data discovery & Governance
Share Metadata of all Data Assets to Data Mesh Catalog
Discover Data Assets via Search
Tag Terms with Data Categories
Support Custom Metadata Attributes
Data and Metadata Profiles (xBOMs): Assemble all relevant information regarding a specific topic (data, assemblies, operation chains, etc.) into a single machine-readable object
Assemble All Relevant Information for Specific Topic into Single Machine-Readable Object
Generate Machine-Readable Asset Inventory Lists for Objects
Perform Dynamic Assembly and Sub-Assembly Management
Advance Analytic Insight Targets
Curate Lineage, Provenance & Pedigree Data from Single Logical Collection
Policy Access Control: Manage access control using metadata for Identity, Role and Attributes for Data Products, Data Objects and Resources with the ability to apply digital policy to achieve precise sharing scenarios
Manage Access Control using Metadata for Identity, Role & Attributes for Data Products/ Objects/ Resources
Apply Digital Policy to Achieve Precise Sharing Scenarios
Apply Discrete Metadata Characterization for Identity, Role & Attribute
Perform Fine Grained Control of Business Rules using PBAC Metadata
Manage Complexity via Focused Rule Creation for Identity, Role & Attribute (Compute & Asset)
Perform Cascaded Rule Evaluation to Optimized Challenge Sequencing Progression from Identity to Role to
Achieve Local Access Control Specifications using Supplemental Characterization Metadata & Rules
Digital Policy Administration: Use logical expressions to describe an unambiguous description of a policy, procedure, or intent, typically using controlled vocabulary with operators to determine a reproducible computed result
Describe Policy/Procedure/Intent via Logical Expressions Using Controlled Vocabulary with Operators to Determine Reproducible Computed Result
Capture Process & Procedures as Computable Expressions using Data, Metadata & Logic Operators
Establish Policy & Procedures using Data, Metadata & Logic Operators
Manage Rule Types Separately (Constraint, Derivation)
Automate Processes & Procedures
Execute Complex Decision Logic at Machine Speed
Execute Complex Logic Scenarios via Chaining/Stacking of Multiple Rules
Mitigate/Correct Error & Risk with Data or Logic Adjustments
Data Exchange Management: Deliver data to a user. This capability should provide delivery in all ways possible like file transfer, database connection, API, cloud storage, access-denied environments, etc.
Determine optimal exchange platform
Verify access via digital policy admin, Policy access control, xBOMs
Data Product Search: Search data products shared on the mesh and allow discovery of data through any vector including full text search
Search Data Products Shared on the Mesh
Full-Text Search for All Data
Search Metadata of All Data Assets
Turn User Query into Query Natural to Each Database
Coalesce All Search Results into Uniform View for User

20

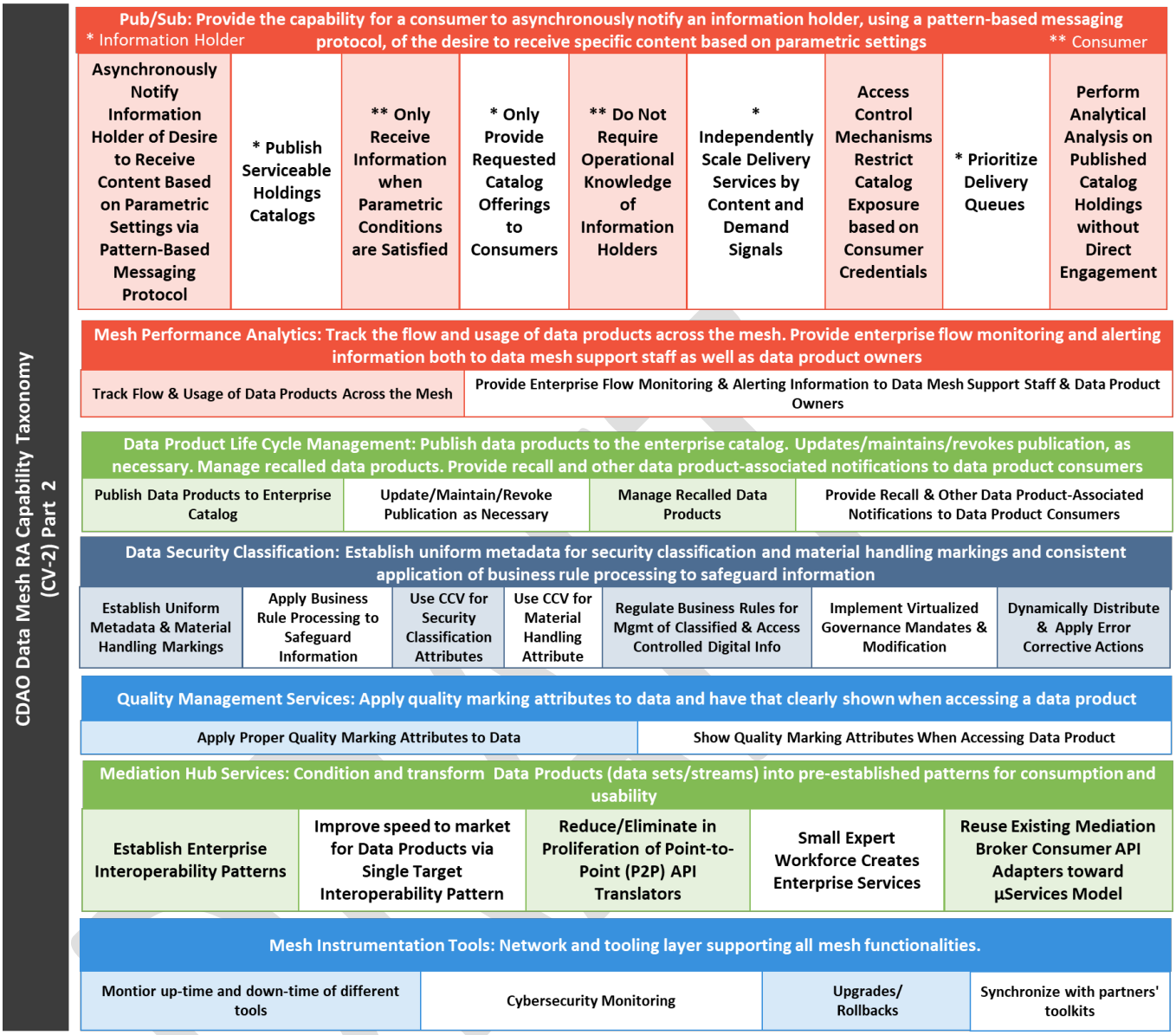


Figure 8: CV-2 Capability Taxonomy

21

Intermediate/Consumer Service / Core/Intermediate Service	Semantic Services **	Data Mesh Catalog **	Data and Metadata Profiles (xBOMs) **	Policy Access Control **	Data Exchange Mgmt **	Data Product Search ***	Data Mesh Pub/Sub **	Mesh Performance Analytics ***	Data Product Life Cycle Mgmt **	Data Security Classification **	Mediation Hub Services ***
UID *	X	X	X	X		X	X		X		X
Semantic Services **	X		X			X				X	
Data Mesh Catalog **						X					
Data and Metadata Profiles (xBOMs) **	X										
Policy Access Control **					X		X				
Digital Policy Administration *				X							
Data Exchange Management **									X		
Data Mesh Pub/Sub **		X									
Data Product Life Cycle Management **		X						X			
Data Security Classification **				X							
Quality Management Services *		X									
Mesh Instrumentation Tools *								X			

Figure 9: CV-4 Capability Dependencies Matrix

The CV-4 Capability Dependencies Matrix in Figure 9 describes the relationships among the mesh service capabilities, defining logical groupings based on the need for those elements to be integrated. The CV-4 provides a means of analyzing the dependencies among mesh service capabilities. The groupings of capabilities are logical with the purpose to guide enterprise service management.

For DMRA, the data mesh service capabilities are grouped into three service categories:

- **Core Mesh Service**: Services originating mesh functions which are utilized by the intermediate mesh services and/or consumer mesh services.
- **Intermediate Mesh Service**: Services that consume other services AND is consumed by other services.
- **Consumer Mesh Service**: Services that consume core and/or intermediate mesh services.

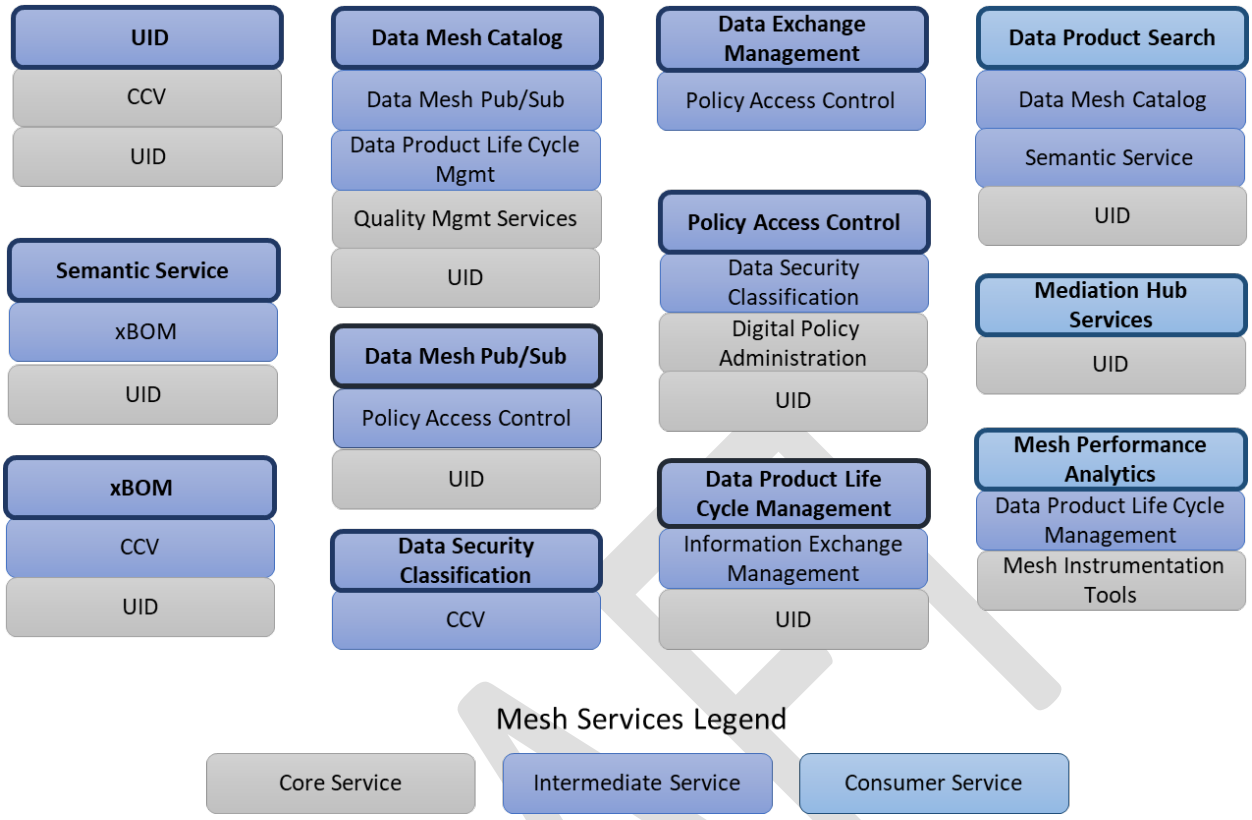


Figure 10: Data Mesh Services Stack

Figure 10 illustrates the primary dependencies of the data mesh services' stacks, where consumer and intermediate services are built on top of their dependent mesh component services to form service assemblies that fulfill end-to-end service capability requirements.

Operational Patterns

The OV-5a Operational Activity Decomposition Tree in Figure 11 describes patterns of operational activities normally conducted while achieving a mission or a business goal. It describes operational activities (or tasks); Input/Output flows between activities, and to/from activities that are outside the scope of the Architectural Description. This ties the DMRA to strategic goals defined in the DoD Data, Analytics, and AI Adoption Strategy through roles and responsibilities. It facilitates operational planning, assists in Information flow analysis, and supports task analysis to determine training needs.

The solid-block horizontals in the OV-5a isolate the functional purposes of the mesh. These include improving the foundational data quality management, delivering capabilities to boardroom and battlefield, strengthening governance, and removing policy barriers, advancing the data, analytics, and AI ecosystem, and expanding digital talent management.

23

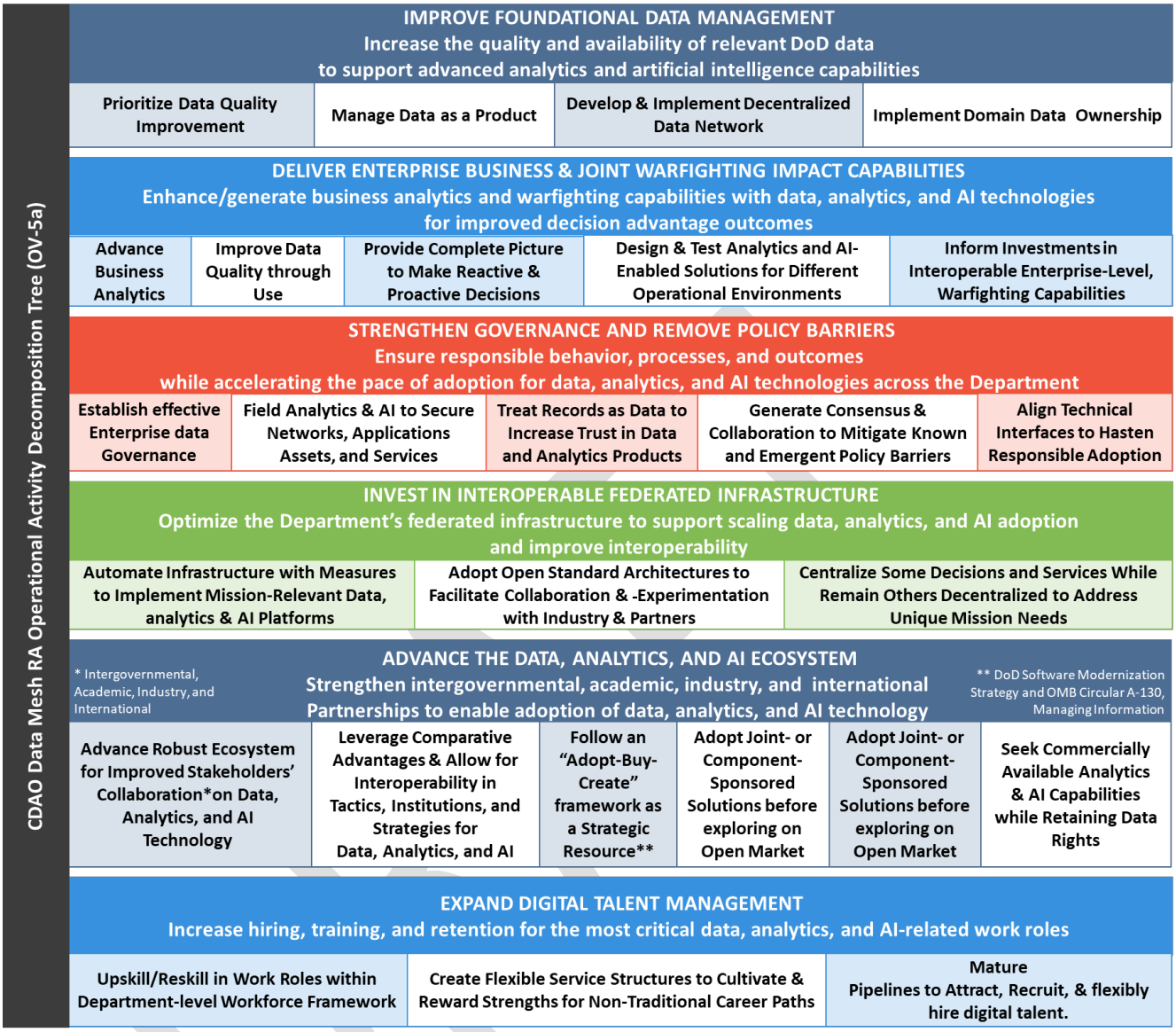


Figure 11: Operational Activity Decomposition Tree (OV-5a)

Service Patterns

The Services Viewpoint (SvcV) SvcV-1 Services Context Description in Figure 12 addresses the composition and interaction of Services within the Data Mesh. It incorporates human elements as types of Performers, Organizations, and Personnel Types. This DoDAF diagram defines service concepts or options, service resource flow requirements capture, and service integration management by illustrating how the human aspects of the mesh and interact with Services.

24

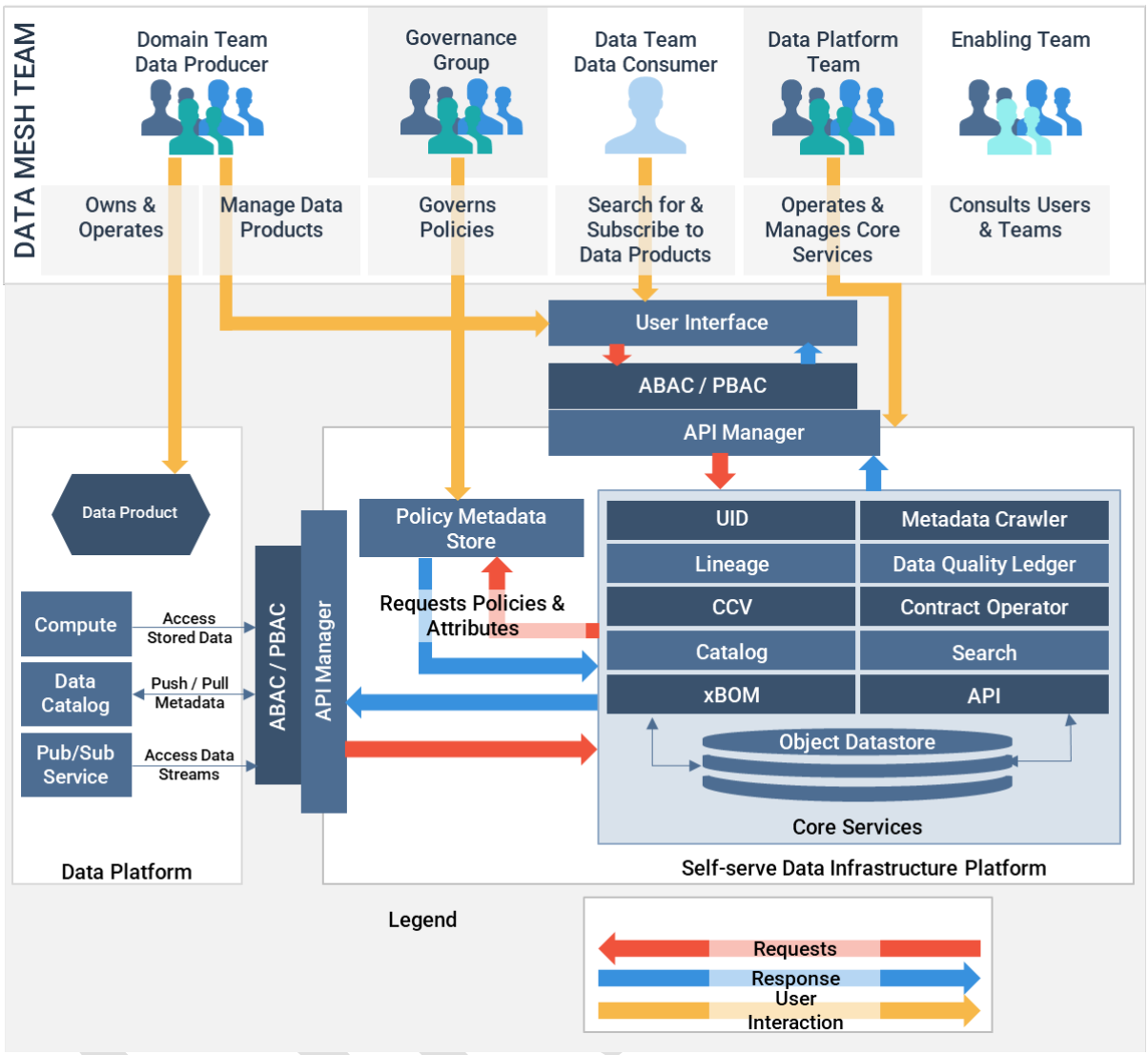


Figure 12 Data Mesh Services Context Description (SvcV-1)

5. Use Case Walkthroughs

High Level Use Case

The OV-6c Event-Trace Description in Figure 13 provides a time-ordered examination of the resource flows in a generic scenario, tracing the lifecycle of a notional question, regardless of type through the mesh. With the exception of the monitoring services, it indicates when each of the proposed mesh capabilities is triggered. When walking through this scenario remember that the actor may be human or machine.

25

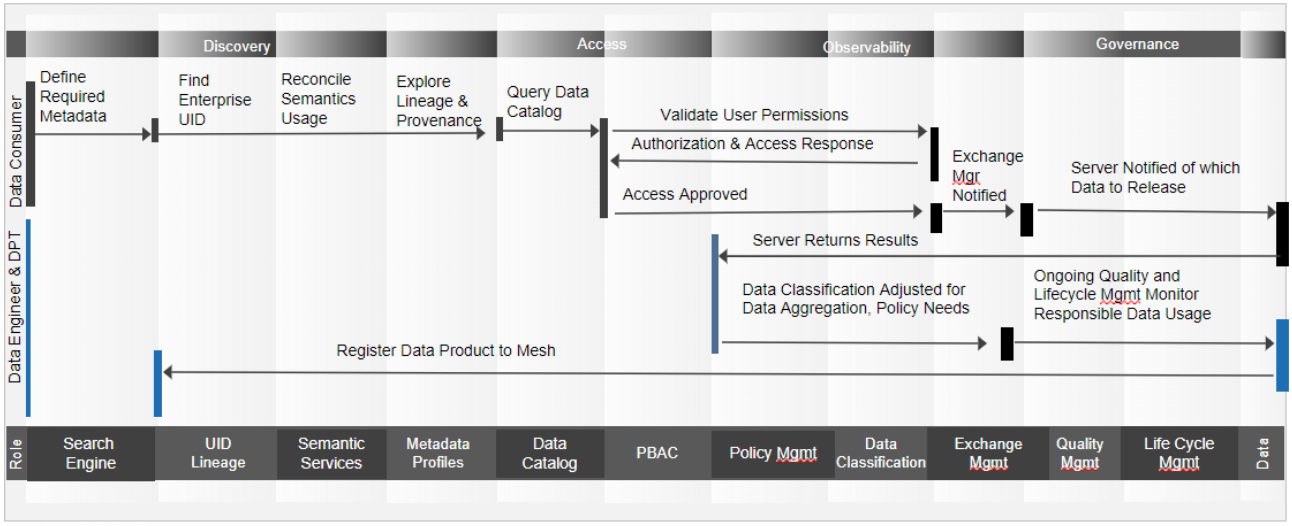


Figure 13: Data Mesh RA Event-Trace Description (OV-6c)

In the upper left, the Data Consumer (DC) is posed a question (not pictured), causing them to query the mesh for what data or data products are available, utilizing the Search Engine service. The query returns a variety of data that it believes best matches the DC's needs, but these must be made comprehensible to the DC's local understanding as they are tagged with conflicting identifiers and expressed in different semantic standards. UID Lineage and Semantic Services perform this translation back to the DC's local standards. These two services concurrently provide lineage and provenance of the resulting data by consulting the Metadata Profiles, thus providing insight into any transformations, origin, or other permutations were applied to the data so the DC can determine which data is best for their use case.

Now the DC may query the Data Catalog since they know which data they want to use or at least explore further. This naturally requires permissions, so triggers a validation of the DC's credentials according to Policy Access Control, Data Classification, and Policy Administration. This triggers the Data Exchange Manager (e.g., Application Protocol Interface [API], cloud storage solutions, and access-denied environments) to transfer data. As the data is being transferred, the Quality Management and Life Cycle Management services are updated to both check that the data transferred matches the request and to register that it was requested for meta-analytics on mesh usage.

Assuming that the query returns data from multiple sources, data classification must be adjusted for aggregation and policy requirements. It then re-engages Quality Management and Life Cycle Management to ensure these activities are known to the mesh and are returning data as was originally requested with little to no quality degradation. Finally, after the DC has performed the required analysis they register that product back to the Search Engine to be available for future searches.

This representative flow of a human, digital workflow, or machine movement through the mesh as it pursues initiatives set forth by the Mission Areas, Capability Application, and Capability Development constructs introduced in Figure 1 encourages the reader to imagine any other use case.

Battlefield Actor Use Case

The Battlefield Actor is the warfighter, commander, in-theater logistician, and "Boots on the Ground" (BOG) actor. Typically, Battlefield Actors are engaged in field operations, and, as a reminder, may be both

human and machine. They mediate real-world tactical execution with policy and governance requirements.

Figure 14 provides a notional scenario of the resource flows among the 15 proposed mesh components using the warfighting scenario of a military exercise in which two friendly aircraft (I.e., in military terminology they are said to be flying "blue forces") and one friendly ship identify an enemy aircraft (I.e., "red forces"). Concurrently, an ocean-worthy civilian sailboat is in the open ocean (I.e., in seafaring terminology it is called a "blue water vessel") within range, triggering safety protocols. Beginning with the upper left:

A) The aircraft and ship use (1) Unique Identifier (UID) component to identify each other as friendly or not associated. Off to the right, this verification requires that each entity is granted access via the peripheral component of (5) Policy Access Control. *Result: The friendly platforms know who is friendly vs. Enemy vs. Civilian bystander.*
B) In communications, friendly aircraft state they are "flying blue forces" and the ship indicates the presence of a civilian "blue-water vessel" nearby. Recognizing they are defining "blue" differently, they each employ (2) Canonical Controlled Vocabulary (CCV) mesh component to translate without requiring human intervention. This action initiatives engagement with (14) Mediation Hub Services, which will be used by the next few actions. (14) is an orchestrator of sorts, enabling the components to be used by human-to-human, human-to-machine, and machine-to-machine interactions. *Result: friendly forces are able to express themselves in their relative lexicons, lowering cognitive requirement of having a native understanding of each other's languages.*
C) Preparing to approach the enemy aircraft, the friendly forces query the (3) Data Catalog to know each other's available firepower, ensuring they are ready for what could ensue. Since (3) is a static data catalog, it relies on the (8) Data Product Search component, which interacts with (1) and (2) to make it easier to search the data catalog with minimal to zero manual work. *Result: parties now know what they have available, thus may call for backup if necessary.*
D) This step has two parts:
    a. The ship asks friendly aircraft to describe the enemy aircraft, information which is digitally stored in (4) Data and Metadata Profiles or [x]BOMs. This pings the neighboring component (13) Data Quality Management to ensure that information retrieved is, for instance, from the most recent [x]BOM and not an archived version. *Result: parties are now aware of the nature of the enemy aircraft, including its make and model, number of pilots, its peer US and allies model, and country of origin, etc.*
    b. Since the nature of the enemy is within the domain more natural to the aircraft participants, the ship defers to air support to provide protocols of engagement with the aircraft. This critical step utilizes both (6) Digital Policy Administration to ensure, for instance, that policies produced align to the current mission and (13) Data Security Classification as some mission sets or enemy aircraft interactions may be at classification levels exceeding the current mission. *Result: friendly forces have a clear path forward on how to execute the mission. If necessary, they will call in support with appropriate classifications.*
E) Steps A-D constitute the operations of the Data Mesh (DM) proper and leverage (7) Data Exchange Management (e.g., APIs [Application Programming Interface], S3 buckets, radio frequency [RF]) to exchange information. The perimeter of the DM proper runs both (10) Mesh

27

Performance Analytics to monitor the health of the mesh and (11) Data Product Life Cycle Management, which, for example, manages the permissions and other considerations for the new data submission that this exercise is generating as it potentially generates new intel from interacting with the enemy aircraft. *Result: data exchange, querying, and updates are performed safely and within the requirements of the DM.*

F) All elements to date are encompassed inside (15) Mesh Instrumentation Tools, which optimizes performance of the mesh and prioritizes data based on behavior analytic data streams. *Result: the system is at least partially self-healing since it is able to detect and quickly diagnose issues as they arise in production, while highlighting redundancies in the system-to-system administrators.*

G) Departing the data mesh proper, (7) Data Exchange Management passes findings to the decision makers, in this case the pilots and ship captain, to determine whether to fire. After firing the missile, via (3) Data Catalog, the mesh understands that the ship needs to re-up its missile inventory upon return to port so it employs (9) Data Mesh Pub/Sub to send an asynchronous message to logistics to be prepared with a backfill missile.

Figure 14 develops a concrete understanding of how the mesh supports the DoD mission space using the example of a military exercise.

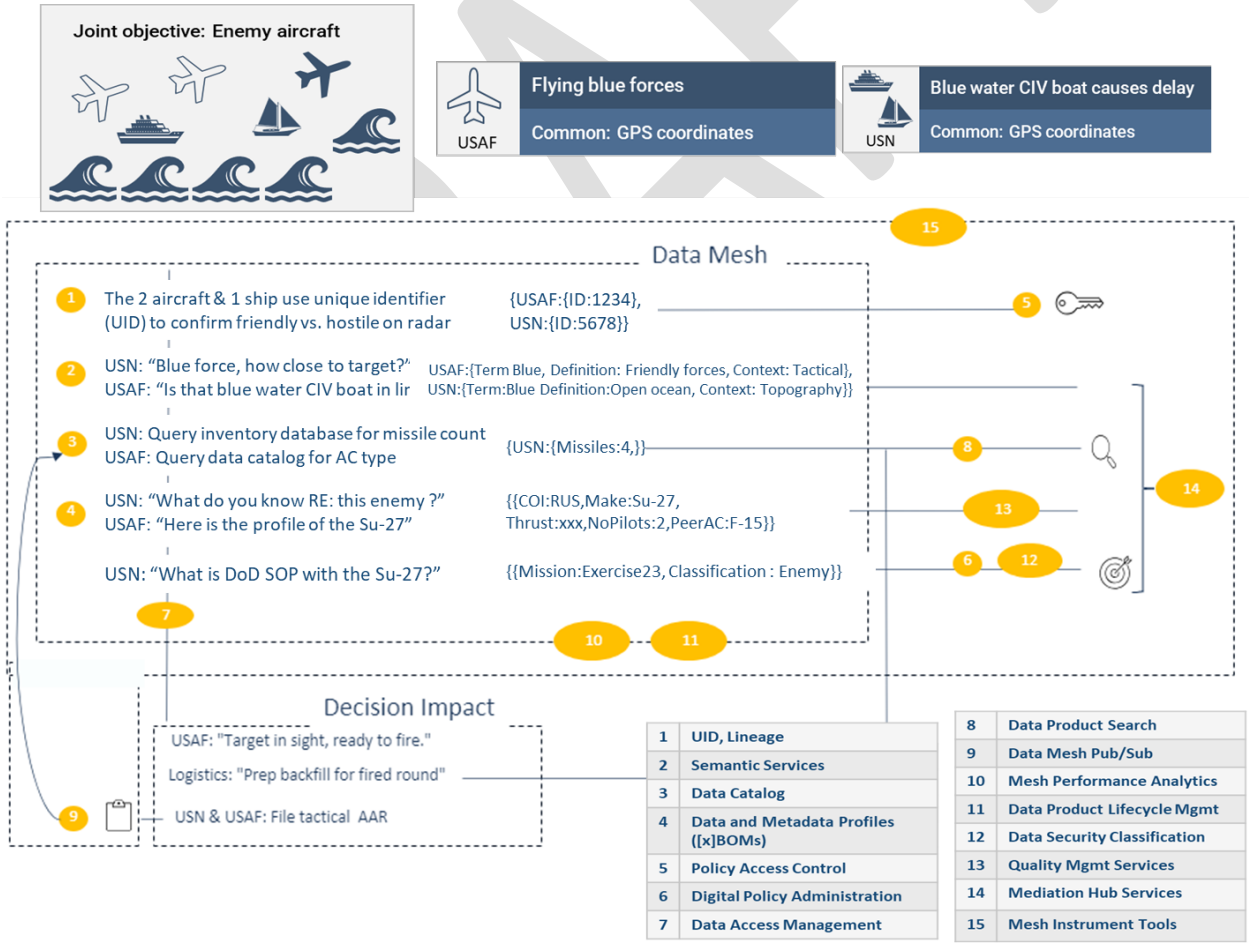


Figure 14: Battlefield Actor DMRA Use Case

28

6. Summary

This document has examined the architecture of the CDAO Data Mesh from multiple perspectives to develop a high-level understanding of how the mesh will support the Department's need for seamless access to federated data, analytics, and infrastructure for digital transformation and Artificial Intelligence (AI). It has provided a strategic, principled and technically grounded approach to enabling decentralized data management across DoD and among partners by proposing that the core capabilities include privacy, classification, access management, and continuous monitoring.

It fulfilled its purpose to guide and constrain instantiations of the mesh (Solutions Architectures) by providing frameworks and guidance for SAs to use in order to ensure compatibility with the enterprise mesh. These frameworks include:

- Common language for the various stakeholders
- Connection points among solutions' implementations
- Ability to validate SAs with the enterprise RA
- Enables adherence to common patterns versus strongly dictating standards
- Use case walkthroughs illustrating usability

Appendices

Appendix 1: All -Views 1 (AV-1) Overview and Summary Information

The AV-1 is an executive-level summary of the CDAO DMRA. This release of the AV-1 is intended to focus on the architecture development effort by documenting the scope of planned RA. This product will be amended and released again later in the architecture development cycle.
Architecture Project Identification

<table>
<tr><td>Name</td><td>CDAO Data Mesh Reference Architecture (DMRA) 1.0</td></tr>
<tr><td>Architect</td><td>CDAO Chief Enterprise Architect</td></tr>
<tr><td>Developed By</td><td>CDAO</td></tr>
<tr><td>Assumptions and Constraints</td><td>• Key stakeholders and subject matter experts will provide information to the CDAO RA team and will review and validate architecture development deliverables.
• CDAO will coordinate RA development efforts with the stakeholders to ensure that they are integrated and do not have overlapping scopes.</td></tr>
<tr><td>Approval Authority</td><td>CDAO Chief Technology Officer</td></tr>
<tr><td>Date Completed</td><td>Version 1: 31 Jan 2024. Version 2: 28 Feb 2024</td></tr>
<tr><td>LOE and Development Costs</td><td>Level of effort, projected and actual costs to develop the Data Mesh RA may be requested from CDAO.</td></tr>
<tr><td colspan="2">Scope: Architecture View and Products Identification</td></tr>
<tr><td>Products Developed</td><td>• The five key elements of the CDAO DMRA are described below:
o Models will be created in Department of Defense Architecture Framework (DODAF) format.
▪ Note: Additional models may be created during the assessment to facilitate leadership decision making. These models will be added to the list of products developed.
<table>
<tr><td>Purpose: Introduction, overview, context, scope, goals, purpose, why needed, and when and how used</td><td>• AV-1 Overview & Summary Information
• CV-1: Vision – overall strategic concept and high-level scope
• OV-1 High Level Operational Concept Graphic – executive operational summary level of what solution architectures are intended to do and how they are supposed to do it</td></tr>
<tr><td>Principles: Foundational organizational rules, culture, and values that drive technical positions and patterns</td><td>• OV-6a Operational Rules Model
• OV-6b Operational State Description
• SvcV-10a Services Rules Model
• SV-10a Systems Rules Model
• OV-4 Organizational Relationships Chart – architectural stakeholders</td></tr>
</table></td></tr>
</table>

<table>
<tr><td></td><td>
<table>
<tr><td>Technical Positions & Policies</td><td>• StdV-1 Standards Profile – standards, specifications, guidance, policy applying to elements of the solution architectures.</td></tr>
<tr><td>Architectural Patterns:

• Generalized patterns of activities, service functionality and system functionality and their resources, providers and information/ data resource flows

• Generalized scenario patterns of sequenced (sequential concurrent) responses by activities, services and system functions (together with their resources) to synchronous/ asynchronous timed events</td><td>
<table>
<tr><td>Operational Patterns

• OV-2 (multiple) Operational Resource Flows
• OV-5 {a, b} Activity diagrams
• OV-6c Event-Trace Description

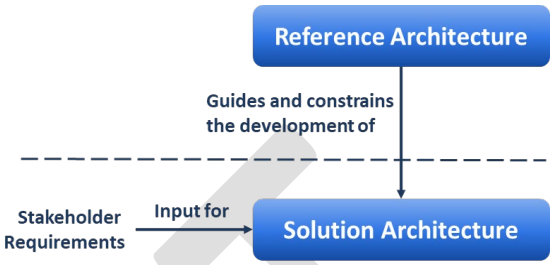
Service Patterns

• SvcV-1 (multiple) Service Interfaces *
• SvcV-2 Service Resource Flows *
• SvcV-4 Service Functionality
• SvcV-10b Service State Transitions
• SvcV-10c Services Event-Trace</td><td>System Patterns

• SV-1 (multiple) System Interfaces
• SV-2 System Resource Flows
• SV-4 System Functionality
• SV-10b System State Transitions
• SV-10c Systems Event-Trace Description</td></tr>
<tr><td colspan="2">Event-Based Scenario Patterns of Dynamic Behavior

• OV-6c Event-Trace Description
• SvcV-10c Services Event-Trace Description
• SV-10c Systems Event-Trace Description</td></tr>
</table>
</td></tr>
<tr><td>Vocabulary</td><td>AV-2 Integrated Dictionary- definitions of terms used throughout solution architectures</td></tr>
</table>
</td></tr>
<tr><td>Time Frames Addressed</td><td>The CDAO Data Mesh RA will refer to the year 2023 and out.</td></tr>
<tr><td>Organizations within the Scope</td><td>The CDAO Data Mesh RA will apply to ADVANA, Longbow (JADC2 DIL, UNACORN), Alpha-1, and SUNet programs at the time of development.</td></tr>
<tr><td colspan="2">Purpose and Viewpoint</td></tr>
<tr><td>Purpose</td><td>The primary purpose of the CDAO Data Mesh Reference Architecture (RA) is to provide a blueprint to guide and constrain the instantiations of solution architectures. The RA is considered an organizational asset in:</td></tr>
</table>

31

	• Providing common language for the various stakeholders • Providing consistency of implementation of technology to solve problems. • Supporting the validation of solutions against proven RAs. • Encouraging adherence to common standards, specifications, and patterns.  The DoDAF defines Solution Architecture as a framework or structure that portrays the relationships among all the elements of something that answers a problem. It describes the fundamental organization of a system, embodied in its components, their relationships with each other and the environment, and the principles governing its design and evolution. Solution architecture instantiations are guided and constrained by all or part of a RA where the generalized and logical abstract elements of the RA are replaced by real world, physical elements according to the specified rules, principles, standards and specifications. The derived DoD definition for RA is: "Reference Architecture is an authoritative source of information about a specific subject area that guides and constrains the instantiations of multiple architectures and solutions."
Questions to be Answered	• How to invest in interoperable, federated infrastructure? • How to advance the data, analytics, and AI ecosystem? • What organizational changes are needed to accommodate a self-serve data platform? • What use cases can be described to show the value of a self-serve data platform? • How are the use cases motivating the services needed to enable a serve data platform?
Architecture Viewpoint	The CDAO DMRA is developed from an enterprise-level perspective in support of the Department's unified approach across data, analytics, and AI activities, using an agile strategic approach that guides decentralized data management action across DoD, as outlined in the DoD Data, Analytics, and Artificial Intelligence Adoption Strategy, to accelerate and scale decision advantage outcomes towards the Department's digital transformation goals.

	• Superior battlespace awareness and understanding • Adaptive force planning and application • Fast, precise, and resilient kill chains • Resilient sustainment support • Efficient enterprise business operations
Context	
Mission	The Department will continuously seize opportunities presented by iterative technology advancements, at the speed of relevance and at the scale of our global mission to advance digital transformation. To do so, the Department requires a unified approach across data, analytics, and AI activities; an educated, empowered workforce skilled at incorporating commercial teams and tools; continued advanced research and rapid experimentation; and effective integration with our Allies and partners.
Vision	The Department as a data-centric organization that can employ data supporting advanced capabilities for operational advantage, increased efficiency, and oriented enterprise data management activities toward the VAULTIS goals to make data: *Visible* – Consumers can locate the needed data. *Accessible* – Consumers can retrieve the data. *Understandable* – Consumers can recognize the content, context, and applicability. *Linked* – Consumers can exploit data elements through innate relationships. *Trustworthy* – Consumers can be confident in all aspects of data for decision-making. *Interoperable* – Consumers have a common representation/ comprehension of data. *Secure* – Consumer data is protected from unauthorized use/manipulation.
Goals for RA Development Effort	• Provide useful guidance and direction on the development and use of RA as a tool to guide and constrain architecture and solution development. • Provide common language for the various stakeholders. • Provide consistency of implementation of technology to solve problems. • Support the validation of solutions against approved RAs. • Encourage adherence to common standards, specifications, and patterns. • Guide and constrain solution architecture instantiations where the generalized and logical abstract elements of the RA are replaced by real world, physical elements according to the specified rules, principles, standards, and specifications. • Serve as a reference foundation for architectures and solutions and may also be used for comparison and alignment purposes.

33

Rules, Criteria, and Conventions Followed	Rules: • Architecture products shall be developed and decomposed only to the level of detail required to adequately portray enterprise "To-Be" business capability improvements and transformation priorities. • The Architecture Support will adhere to the DoDAF v2.02, where needed. • The Architecture Support will integrate and maintain architecture products in one integrated repository, using one methodology and one consistent notation. • The Architecture Support will depict architecture products in sufficient detail to provide an unambiguous interpretation of enterprise-level business and technical strategies. Criteria and Conventions – Guidance contained in the DoD Business Enterprise Architecture (BEA) Development Methodology (BDM) and Architecture Product Guide (APG) provide the criteria and conventions followed during the CDAO Data Mesh RA development for methodology, content, and format.
Tools and File Formats to be Used	
MS Word, MS Excel, MS PowerPoint, MS Visio	

34

CITATIONS

[1] U.S. Department of Defense Chief Information, "The DoDAF Architecture Framework Version 2.02," August 2010. [Online]. Available: https://dodcio.defense.gov/Library/DoD-Architecture-Framework/

[2] U.S. Department of Defense, "Data, Analytics, and Artificial Intelligence Adoption Strategy - Accelerating Decision Advantage," U.S. Department of Defense, 2023.

[3] Z. Dehghani, Data Mesh: Delivery Data-Driven Value at Scale, O'Reilly Media, 2022.

[4] Office of the Assistant Secretary of Defense, "Reference Architecture Description," 2010. [Online]. Available: https://dodcio.defense.gov/Portals/0/Documents/Ref_Archi_Description_Final_v1_18Jun10.pdf

[5] CDAO Data Mesh Strategy.

[6] "How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh," martinfowler.com, 20 May 2019. [Online]. Available: https://martinfowler.com/articles/data-monolith-to-mesh.html