Develop an Intelligent Data Strategy for Defense Agencies with Data Cataloging

Accelerate Data Discovery and Access
for Operational Advantage

Meet Minimum Metadata Requirements for a Federated Data Catalog

To ensure mission success, the United States Department of Defense (DoD) is focused on improving data management across all of its components. The DoD data strategy framework outlines the key characteristics, requirements, and capabilities needed to achieve an enterprise approach to data management (see Figure 1). Building on this framework, a recent memorandum for chief data officers details the DoD's federated data catalog model.
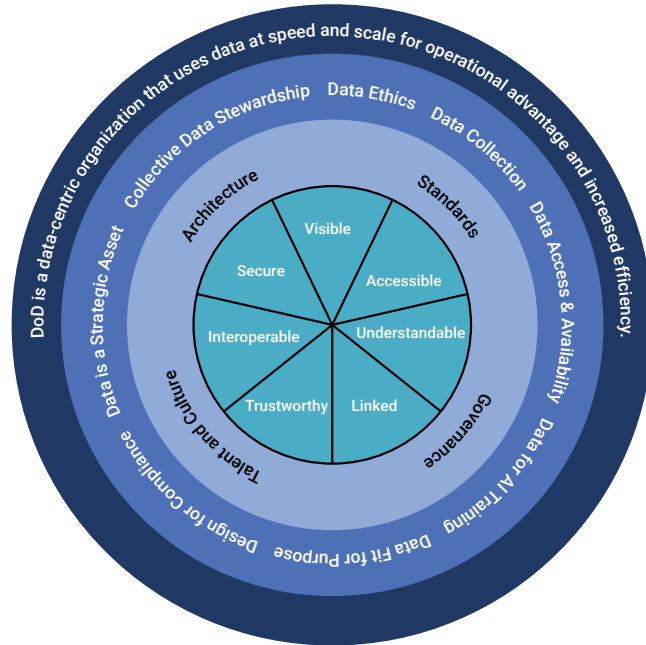

Figure 1. DoD Data Strategy Framework

To meet the DoD's goal of making data visible and enabling consistent records of its assets within the federated data catalog, your agency must adopt the right enterprise data cataloging technology.

Key Capabilities for Data Cataloging	
Capability:	**Allows data users to:**
Visible	Locate the needed data
Accessible	Retrieve data
Understandable	Find descriptions of data to recognize the content, context, and applicability
Linked	Exploit complementary data elements through innate relationships
Trustworthy	Be confident in all aspects of data for decision-making
Interoperable	Share a common representation and comprehension of data with data producers
Secure	Protect data from unauthorized use and manipulation

Understand the Value of Data Cataloging and Governance

An enterprise data catalog acts as a central place to collect, index, relate, annotate, and share knowledge about agency data assets. It can support mission success by serving as:

• The **foundation** for your digital transformation journey
• The **source** for making data-driven decisions
• A **single knowledge repository** that enables innovation and collaboration through data discovery and exploration

What's more, a data catalog can accelerate the use of your data as a business and mission asset.

Why Should Commands Adopt Data Catalog Technology?

Warfighters and their enabling agencies need to find and use data in diverse ways. To keep pace with changing conditions and events, your users need the ability to rapidly understand and marshal data assets.

Self-service features allow data users across the organization to quickly explore, discover, sanitize, and reuse data. As they see how a data catalog can help them achieve results faster and more accurately, users understand and embrace the value of adopting a data catalog.

"We're going to make sure that we have the data we need, that we can consistently deliver the data that's required, so that the warfighter can count on that model to be there like they can count on their battle buddy when they show up and they're ready to hit an objective."

-   David Spirk, Chief Data Officer, U.S. Department of Defense, closing VIP keynote, Data In Action Summit 2020

How Can You Ensure the Successful Adoption of Data Catalog Technology?

Take the following steps to help your agency streamline your adoption of a data catalog solution:
• **Formulate** your program strategy.
• **Start with a pilot**, prioritizing a few key data sources with low complexity and high user impact.
• Load and **enrich the data catalog** with relevant content.
• Train, **engage, and communicate often**.
• Obtain feedback from users and **apply lessons learned** before expanding to the next wave.

Build a Data Catalog Based on Proven Best Practices

There are many paths to formulating and implementing your data catalog program strategy.
Following are three recommended steps.

Formulate Your Program Strategy

Define a data catalog strategy that:
• Aligns and supports your agency vision
• Focuses on delivering business impact with use cases linked to business drivers
• Addresses user pain points
• Communicates program value and long-term direction

Implement Your Program Strategy

Remember that you'll need to continuously refine and expand your data catalog over time. The most successful agencies follow these steps:
• Start small with a pilot project
• Ingest and enrich metadata
• Assign users and responsibilities
• Train users, track usage, and share insights

Enrich, Refine, and Expand Your Data Catalog

As your data catalog matures, you will need to enrich it to meet your agency's evolving needs. We recommend that you begin by enriching assets with:
• Business title and description
• Business terms
• Data owner and data steward
• Certification of data assets
• Data profiles
• Custom attributes such as business usage and category
• Data domains, such as personally identifiable information, protected health information, and any domains customized for your agency
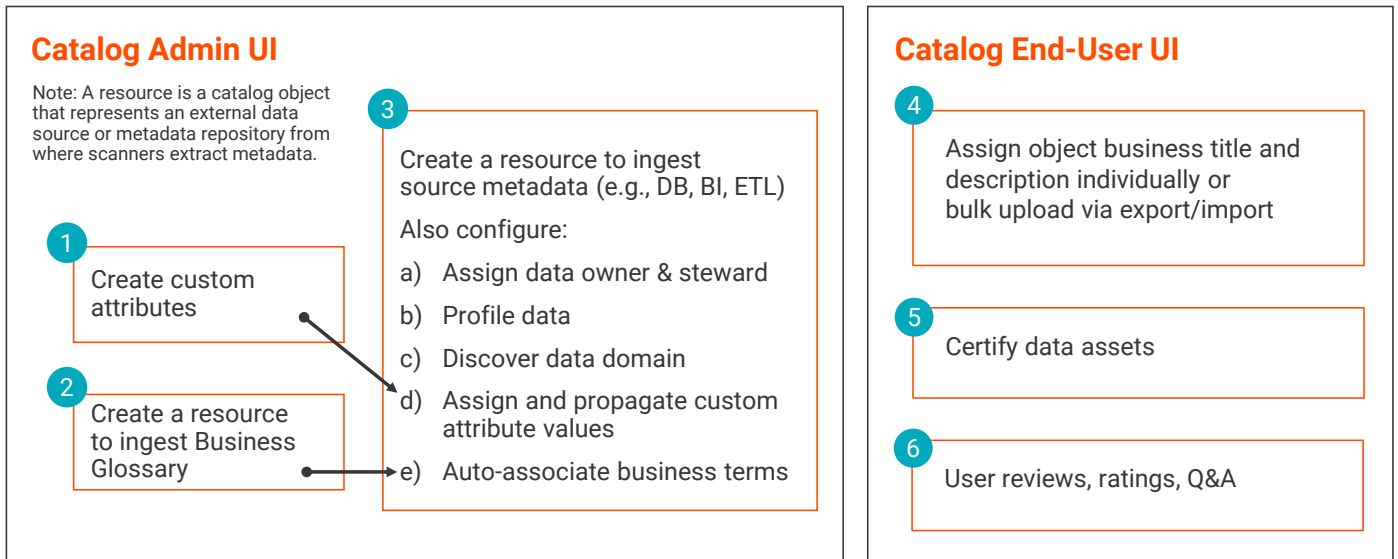
Figure 2 illustrates a data enrichment approach.

Catalog Admin UI

Note: A resource is a catalog object that represents an external data source or metadata repository from where scanners extract metadata.

1 Create custom attributes

2 Create a resource to ingest Business Glossary

3 Create a resource to ingest source metadata (e.g., DB, BI, ETL)

Also configure:

a) Assign data owner & steward

b) Profile data

c) Discover data domain

d) Assign and propagate custom attribute values

e) Auto-associate business terms

Catalog End-User UI

4 Assign object business title and description individually or bulk upload via export/import

5 Certify data assets

6 User reviews, ratings, Q&A

Figure 2. Practical Steps for Enriching a Data Catalog

Extract Value from Your Data with AI-Powered Data Discovery

AI-powered data catalog solutions from Informatica can help you extend your understanding of data assets while improving confidence and trust in data-driven business insights. Our intelligent data catalog solutions support AI-powered, metadata-driven discovery, inventory, and preparation—giving your data assets new meaning and relevance. By scanning and cataloging enterprise metadata at scale, these intelligent, automated solutions provide context and organization for your data.

Key Characteristics of Automated Data Discovery

Informatica offers AI-powered automated data discovery solutions with the following characteristics.

Intelligent data classification: Our solutions help you classify data fields by applying semantic tags (that is, data domains) to columns. They automatically apply the same tag to similar columns in different tables. Once data domains are discovered, solutions can assemble individual fields into higher-level business entities.

Intelligent, semantic search: Automated data discovery powered by advanced machine learning algorithms and natural language processing enables Google-like semantic search. This characteristic dramatically speeds the process of finding data across diverse, distributed enterprise systems.

Data profiling: By identifying and understanding characteristics such as data value frequencies, value distribution, and anomalies and missing values, our solutions help you identify data quality issues and increase data accuracy across the organization. Profiling statistics can also be used to identify similar datasets and data relationships.

Automated, end-to-end lineage: Comprehensive data lineage makes it easy to understand where data originates and how it changes over the entire data lifecycle. Advanced relationship discovery features can identify relationships, such as joins and primary keys, by inferring joins across datasets at scale. This insight helps you understand how a change will potentially impact users, business processes, and reports.

Key Use Cases for Automated Data Discovery

AI-powered data discovery solutions from Informatica lay the foundation for your critical digital transformation initiatives.

Advanced Analytics – Enables data scientists and DataOps teams to leverage AI and machine learning (ML) and rapidly find the data they need. In iterative ML projects that require additional data assets, AI-powered data discovery can help DataOps teams find trusted datasets to make available in data pipelines and streamline data preparation.

Customer 360 – Helps you build the data foundation for successful customer engagement by discovering and cataloging customer data wherever it resides, identifying and classifying customer data domains, profiling data quality, and providing visibility into data lineage across customer touchpoints.

Modernizing Data Warehouses and Data Lakes in the Cloud – Delivers end-to-end visibility and lineage across your environment so you can perform detailed impact analysis across data assets, resources, and users.

Data Governance and Privacy – Helps you discover and classify data—including sensitive data such as personal information —so it can be trusted, understood, and secured. By automatically relating data to business terms and definitions, our intelligent, automated data discovery solutions can provide business context to data for enterprise data governance and fuel business value with trusted data.

Meet Your Mission with Informatica Enterprise Data Management Solutions

Only Informatica offers a comprehensive selection of data management solutions that help transform the Defense Department into a data-centric organization (see Figure 3). Enterprise data management solutions from Informatica help your agency manage, understand, and responsibly share and protect data in support of their shared mission.
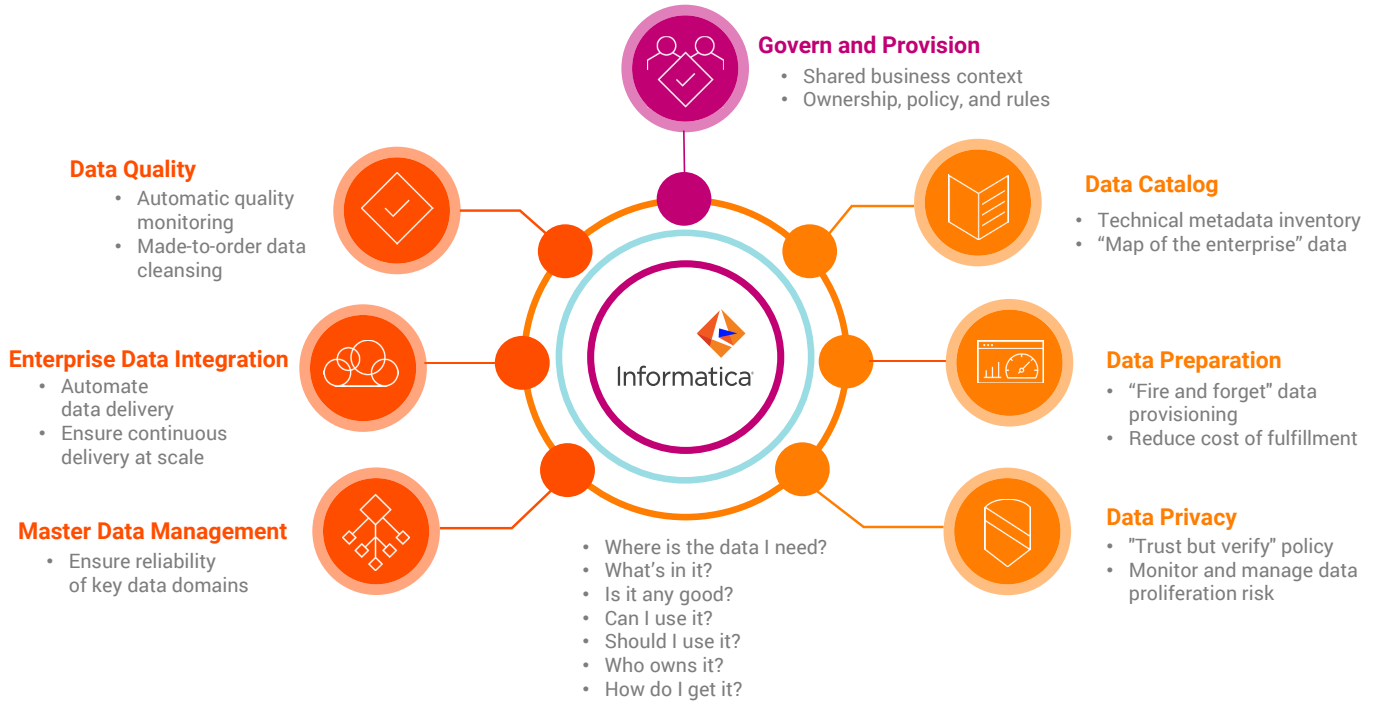

Govern and Provision
- Shared business context
- Ownership, policy, and rules

Data Quality
- Automatic quality monitoring
- Made-to-order data cleansing

Enterprise Data Integration
- Automate data delivery
- Ensure continuous delivery at scale

Master Data Management
- Ensure reliability of key data domains

Data Catalog
- Technical metadata inventory
- "Map of the enterprise" data

Data Preparation
- "Fire and forget" data provisioning
- Reduce cost of fulfillment

Data Privacy
- "Trust but verify" policy
- Monitor and manage data proliferation risk

- Where is the data I need?
- What's in it?
- Is it any good?
- Can I use it?
- Should I use it?
- Who owns it?
- How do I get it?

Figure 3. Informatica Enterprise Data Management Solutions

Learn More

To learn more about how Informatica can support your data catalog requirements, visit www.informatica.com/products/data-catalog