

White Paper

KICKSTARTING YOUR GOVERNANCE ORGANIZATION WITH A DATA CATALOG



Elliot Huebler
Manager



A robust overview from Tredence of driving an enterprise data governance platform and democratizing data by organizing metadata with a data catalog



CONTENTS

01. Introduction	03
02. Defining & Prioritizing Domains	04
03. Creating Roles & Responsibilities	05
04. Domain Interviews & Metadata Collection	06
05. Catalog Standup & Domain Onboarding	09
06. Driving Accountability & Engagement	11
07. Prioritizing & Centralizing Governance Goals	13
08. Summary	14

Introduction

Your journey to establishing data governance has just begun. Perhaps your organization has previously tried to establish a governance program, but it didn't gain traction and uniformity to scale across the enterprise. Regardless, **governance is a necessity for organizations to operate and scale.**

In this guide, Tredence presents an approach to establishing a **centralized governance platform** and **democratizing data across your enterprise** through a data catalog.

A **data catalog** is a tool for collecting and organizing various types of metadata. The process of setting up a robust governance platform is the same as that of creating and organizing the metadata across your enterprise. Following are 5 key governance pillars, in addition to cataloging, for creating or capturing metadata to drive data governance forward.

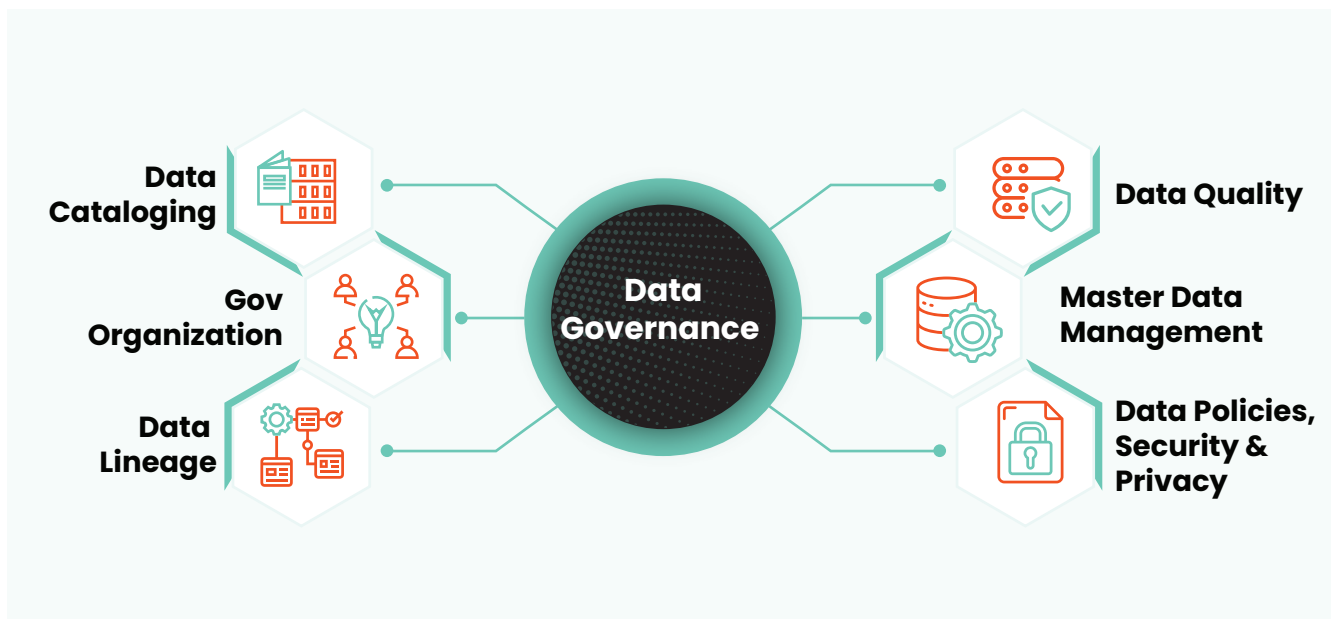


Figure 1: The Critical Pillars of Data Governance

To kick off a governance initiative, we recommend standing up a data catalog and establishing an iterative strategy for onboarding domain users and their metadata as well as onboarding additional governance initiatives. This will allow even a **small governance team** to make a large impact across the enterprise.



Defining & Prioritizing Domains

No matter the size of your enterprise, establishing a simple and iterative process for applying governance to data is a necessity. The recommended approach is to define **data domains** as a structure that users will understand to group data.

They can be defined by functional group (supply chain, recurring revenue, marketing, etc.), product team (product #1, product #2, etc.), region (northeast team, northwest team, etc.), or any other logical grouping. When you eventually drill down into these domains to draw out useful metadata, this then becomes the basis of a top-down strategy for governing data.

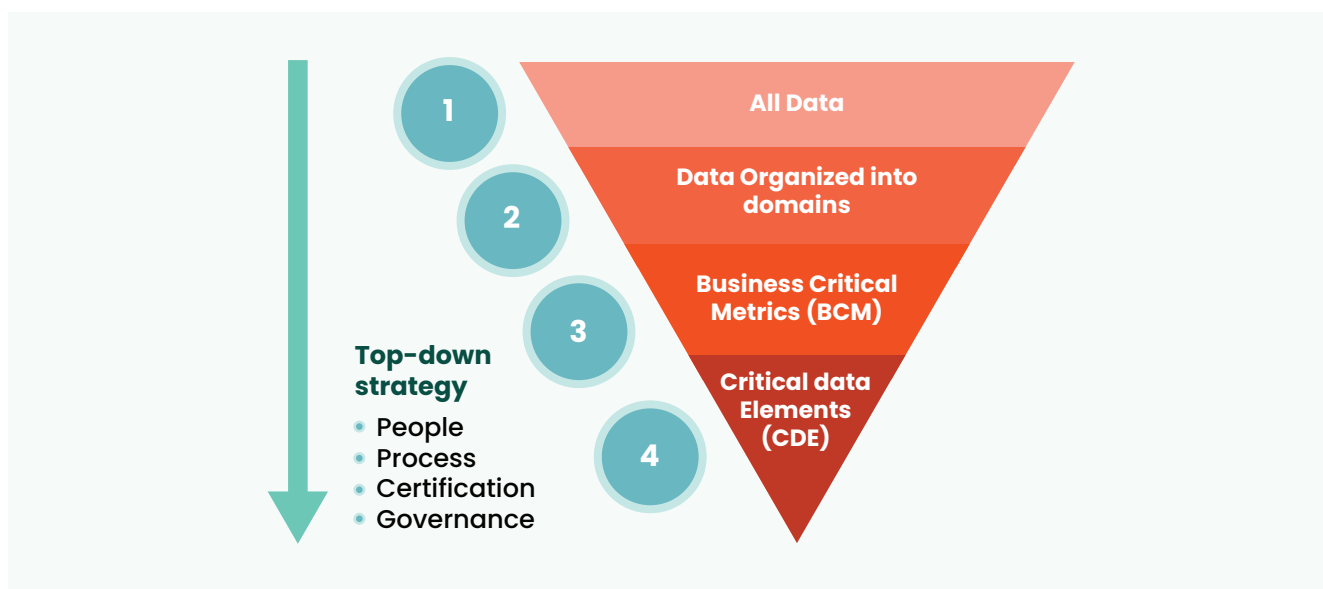


Figure 2: A top-down depiction of the scope of data at the enterprise. The area targeted to be defined is #2.

There will be many domains across your enterprise and the governance team needs to work with leadership to **prioritize data domains** for onboarding into the catalog. There are different ways to prioritize domains, such as:

- 
Criticality to the organization/data that runs the business
- 
Domain's available bandwidth to commit to onboarding
- 
Domain's readiness /availability of metadata for assets

With data domains prioritized, you can select 2 to 5 MVP domains, which will act as the target for metadata collection and onboarding into the catalog. Remember, by harvesting the various metadata attributes from these domains, we are in effect building out the basis of our governance metrics.

Creating Roles & Responsibilities

Shortly, we will conduct interviews across MVP domains to collect metadata. For the sake of efficiency, we can use these interviews to also **set expectations** and assign roles and **responsibilities** for upcoming governance initiatives.



First, we must define roles and assign them at the data object (table/column) level of granularity. These roles span the **5 governance initiatives**; some are outlined in the table below (more can be added to fit your organization's needs).

Governance Area	Role	Responsibility
Governance Organization	Data Steward	Curate basic information about the object: title, description, etc.
	Domain Owner	Curate information about the domain/subdomain and use case for the object
Data Quality	Data Architect	Provide ETL/technical quality rules for the object
	Data Owner	Provide business quality rules for the object
Master Data Management	MDM Expert	Provide hierarchy for object, if applicable
Lineage	Data Architect	Provide upstream source and downstream report/table for object
Security/Privacy/Policy	Policy Owner	Provide PII, data classification, masking policy, policy documents, etc.

Figure 3: A set of example roles and responsibilities spread across governance initiatives

Domain Interviews & Metadata Collection

With MVP domains identified and a scope of roles defined for your governance needs, the next step is to **define the scope of metadata required**. To make interviews with domains go smoother, a **template to clearly set the expectation** of what metadata is being requested from the domains is essential. These templates can then be leveraged later for onboarding into the data catalog.

Metadata may be stored in different ways across domains and different domains may maintain different sets of metadata. It is important to both capture the metadata required for your governance needs and collect additional metadata that domain users will find relevant.



We recommend using two templates to capture metadata.

Template 1: Table Level Metadata

The purpose of this template is to capture metadata across governance initiatives at the table object level of granularity, as seen in the following examples:

	Data Democratization				Security/Privacy/Policy		
Table Location	Title	Description	Steward	Domain	Contains PII	Masking Policy	Data Classification
Database. Schema.Table							
....							

	Data Lineage				Master Data Mgmt.		Data Quality
Table Location	Upstream Table	Downstream Report	Data Architect	Report Owner	Part of Hierarchy	MDM Owner	See Template #2
Database. Schema.Table							
...							

	Any Additional Domain Specific Metadata			
Table Location	Use Case	Links to Knowledge Articles		ID Codes
Database. Schema.Table				
...				

Figure 4: Domain metadata collection templates at the table level of granularity

Template 2: Business Critical Metric Metadata

The purpose of this template is to capture **Business Critical Metrics** (BCMs) and the **Critical Data Elements** (CDEs) that create them. A BCM is a metric used by the domain that is critical in reports, which is derived from a calculation that includes several CDEs or column-level elements.

		Data Democratization					
CDE Location		CDE Title	CDE Description	BCM Title	BCM Description	CDE Owner or Steward	Domain
Database.Schema.Table.Column							
....							

	Data Quality						
CDE Location	DQ Check ID	DQ Check Category	DQ Check Title	DQ Check Description	DQ Check SQL Logic	DQ Business Owner	DQ IT Owner
Database.Schema.Table.Column							
...							

	Data Lineage					
CDE Location	CDE Upstream Location	CDE Upstream Owner	CDE Downstream Location	CDE Downstream Owner	CDE Downstream Report	CDE Downstream Report Owner
Database.Schema.Table.Column						
...						

Figure 5: Domain metadata collection templates at the CDE level of granularity

Note:



We use this template to focus on data quality and lineage metadata. Specifics on building out data quality and lineage implementations and why the templates are structured this way will be discussed in another white paper. Additionally, **other templates may be needed** to capture your organization's specific use cases.

With these templates created, host a meeting with the domain users and use them to set expectations on what metadata you are seeking.

You should **expect missing metadata**, as the domain may not be able to fill in all the metadata you are searching for on the first pass. Iterate with the domain to fill in as much as possible. Remaining gaps may need to be filled in with a third-party tool, structured initiative, or other investment later. Take note of what metadata is missing since it will help when prioritizing future governance initiatives (see Prioritizing & Centralizing Governance Goals).



Catalog Standup & Domain Onboarding

At this phase, a data catalog tool should be selected.

Some critical features to consider in a data catalog to support the goal of data democratization and centralizing governance are:



Simple User Interface to allow new users to quickly find, understand, access, and collaborate on data



Customizability to capture a wide range of cross-domain and cross-governance initiative metadata



Robust API to streamline metadata ingestion from sources such as the templates



Analytics Database to extract important information about userbase engagement and content growth in the catalog
(discussed in *Driving Accountability & Engagement* section)

To **stand up the catalog**, install the platform and create connections to your various databases (catalog vendors often support this phase). With the basic setup complete, the goal is to set up the catalog for MVP domain onboarding. The process of onboarding a domain into the catalog involves enabling **self-service curation and handing off responsibility** from the governance team to the respective domain.

This will allow even a small governance team to work across the enterprise. This onboarding process can be streamlined with the following considerations:

1

Catalog Homepage

A well-designed homepage can help users quickly navigate to data, help articles, and catalog features

4

API Scripts

To quickly ingest/create the metadata for the domains (should all be captured in template)

2

Custom Fields

These fields will capture the domain's specific metadata requirements

5

Reporting Dashboards

To track and audit domain metadata and user engagement

3

Domain Spaces

A location in the catalog where data and knowledge articles can be grouped at the domain level

6

Onboarding Documentation

To teach users about the catalog and define curation expectations

With the above considerations addressed, **domain onboarding** can happen in three simple steps:



Engage with the domain and **set curation expectations** for the catalog. Define these expectations in a document (see the example below).



Ingest the domain's available metadata, including tagging roles and their owners in the catalog through the API.



Begin to track the domain's content coverage and user engagement.

Curation Expectations Example: A simple checklist for curators

1. Curate Domain Spaces

- Add knowledge articles outlining the domain
- Add business context with links, unstructured documents, diagrams, etc.

2. Curate Domain Tables

- Fill in the missing metadata requested in the template
- Tag tables, reports, and other data objects in the catalog to the domain
- Tag as CDE/BCM

3. Author Domain Articles

- Articles outlining domain-specific data information
- Terms or metric definitions (to be grouped into glossary)

Driving Accountability & Engagement

As part of the catalog standup, tools must be built to track and report on the catalog through the lens of **user engagement** and **catalog content** down to the domain level of granularity. From the reports through these lenses, we can derive key metrics on the success of the catalog.



User Engagement	Catalog Content
# of logins	% coverage/growth in coverage of metadata fields
# of visits to tables, articles, domain spaces, etc.	Most visited content
# of metadata fields curated	Content in each domain or assigned to steward
# of articles authored	

Figure 6: Table outlining the KPIs to be monitored in the catalog

These various KPIs will help you understand if domain onboarding has been successful (i.e., that self-service curation is, in fact, happening). They will also help you identify stewards and other data owners who may need to contribute more to the catalog (as well as those who are contributing a lot). Overall, this should help **drive accountability** and **report the catalog's success**.

Thus far, we have focused specifically on the pipeline for onboarding the MVP domains. However, the data catalog is for the entire enterprise, and the door should always be open for any user to join the catalog to discover or contribute. Below are some additional methods to be considered to **drive engagement** and **adoption** of the catalog.










 Survey Users Understand what users like and dislike, address in catalog	 Host Webinars Highlight specific features or use cases in internal webinars	 Understand User Types Is the catalog being used by data or business users?
 Create Badge/Incentive Program Reward users for going through training, curating the catalog, or answering questions	 Make Group Chat for Users Democratize the question asking process with a large collaborative chat space	 Audit Onboarded Domains Identify users who have not logged back, reach out to understand why or give nudge
 Find Data Evangelists Word of mouth advertising from users outside the EDC is powerful, have users tell their stories!	 EDC Newsletter Remind users that EDC is always growing, and new content and features are being added	 Audit Support Tickets Study what users are requesting for help with your data architects or admins

Figure 7: Nine unique ways of driving user engagement

Prioritizing & Centralizing Governance Goals

At this phase we have a fully functioning catalog and methodology to scale and capture metadata across domains. The metadata collected from domains, however, will mainly serve to help democratize data, not satisfy governance initiatives.

To make the data catalog the **central tool for all data governance**, all the related metadata must be routed back to the catalog. Structured projects need to be built out around other initiatives; some examples of critical metadata to add to the catalog are outlined below.

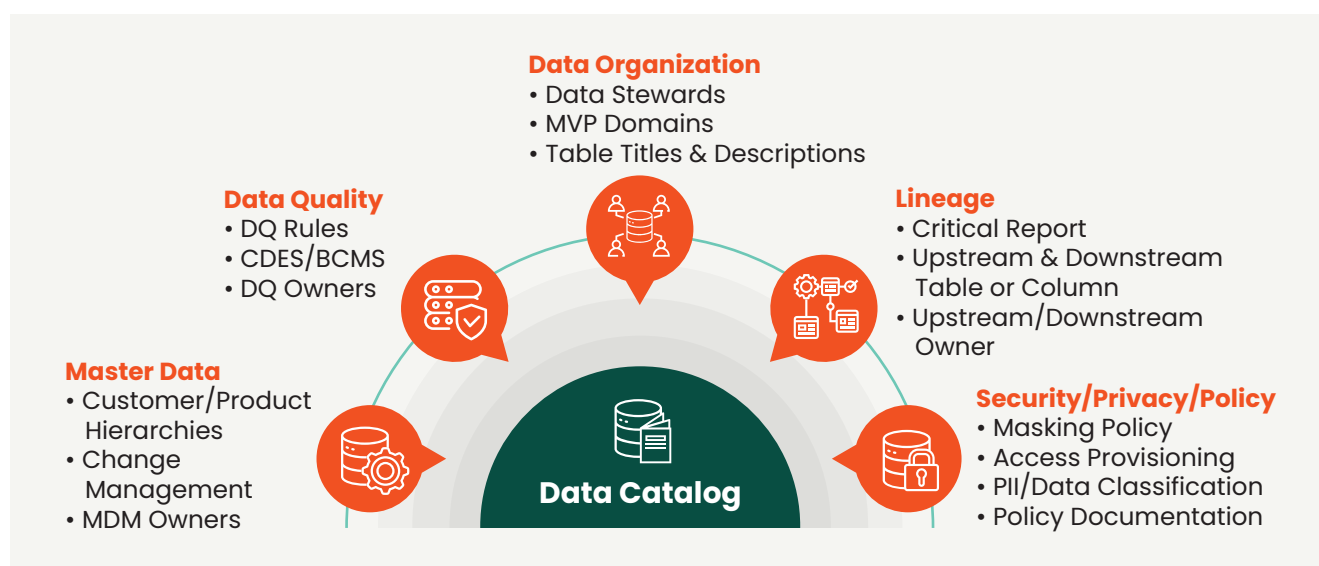


Figure 8: Depiction of cross-governance initiative metadata which routes back to the data catalog

Similar to the data domains in the catalog, **governance initiatives should be prioritized and iterated** with the catalog being the final tool to display the metadata (and therefore report on the success of the initiative).

For each initiative, a third-party tool or custom solution may be required to create metadata and push it back to the catalog (ideally, via the API). These all require different approaches to implementing the solution. To that end, Tredence will be posting several upcoming white papers around the success of these various governance pillars.

Enterprise Data Quality Framework: Proactive & Reactive Approach

Defining and Maintaining Master Data Hierarchies Across an Enterprise

Security & Compliance Considerations for Data Governance

Building a Robust End-to-End Lineage for your Governance Enterprise

Summary

This white paper has outlined how to use a data catalog as a tool to democratize data and centralize governance metadata across your enterprise.

The standup of a data catalog is **preceded by cross-domain prioritization**, metadata collection, and a role assignment initiative, which instructs the initial organization and metadata fields required in the data catalog.



Data domain onboarding consists of **structured conversations** and **the ingestion** of collected metadata. Structured conversations set the expectation of what needs to be curated and helps pass the responsibility from the catalog team to the respective domain. In addition, metadata ingestion is often supported by scripts that interact with the catalog's API.

With the domain onboarded, monitoring scripts/dashboards will report on user engagement and content coverage. Gaps in metadata will be tracked and ideally filled in. Additionally, other engagement pushes outside of the domains will happen, and monitoring scripts will help visualize the success of any engagement drives.

With a domain onboarding pipeline established, start expanding the catalog's governance capabilities by piping back metadata being created in other tools. Finally, create structured projects around any other governance gaps in your organization to connect with the catalog and use it as a tool to report on the project's success.

About Tredence

Tredence is a data science and AI engineering company focused on solving the last mile problem in analytics. The 'last mile' is defined as the gap between insight creation and value realization. Tredence is more than 1,600 employees strong, with offices in Palo Alto, Chicago, Toronto and Bangalore, with the largest companies in CPG, retail, hi-tech, telecom, travel and industrials as clients.

Learn more at: www.tredence.com

Follow us at: [in](#) [🐦](#) [📺](#) [f](#)