

eBook

# Introduction to Data Catalogs



By Dave Wells  
January 2020

Research Sponsored by



This publication may not be reproduced or distributed  
without Eckerson Group's prior permission.



## About the Author



**Dave Wells** is a Senior Analyst leading the Data Management Practice at Eckerson Group, a business intelligence and analytics research and consulting organization. He brings a unique perspective to data management based on five decades of working with data in both technical and business roles. Dave works at the intersection of information management and business management, where real value is derived from data assets. He is an industry analyst, consultant, and educator dedicated to building meaningful and enduring connections throughout the path from data to business value. Knowledge sharing and skills development are Dave's passions, carried out through consulting, speaking, teaching, and writing. He is a continuous learner – fascinated with understanding how we think – and a student and practitioner of systems thinking, critical thinking, design thinking, divergent thinking, and innovation.

## Table of Contents

|   |    |  |    |
|---|----|--|----|
| <b>Introduction</b> .....                                     | 3  | A New Approach to Metadata Management ...                          | 13 |
| <b>Chapter 1: What is a Data Catalog?</b> .....               | 4  | Metadata in the Age of Self-Service .....                          | 13 |
| Starting the Data Cataloging Journey.....                     | 5  | Metadata and the Data Catalog.....                                 | 14 |
| Data Catalog Defined .....                                    | 5  |  |    |
| Why Do We Need a Data Catalog? .....                          | 5  |  |    |
| What Does a Data Catalog Do? .....                            | 6  |  |    |
| <b>Chapter 2: What is Data Curation?</b> .....                | 8  | <b>Chapter 4: Collaboration and Crowdsourcing</b> .....            | 15 |
| More Than Shared Databases .....                              | 9  | Participative and Collaborative Data Management.....               | 16 |
| What is Curation? .....                                       | 9  | Why Collaboration and Crowdsourcing — A Macro View .....           | 16 |
| What is Data Curation?.....                                   | 9  | Why Collaboration and Crowdsourcing — An In-The-Trenches View..... | 16 |
| Who Are the Data Curators? .....                              | 9  |  |    |
| What About Data Stewards?.....                                | 10 | <b>Chapter 5: Driving Data Catalog Adoption</b> .....              | 18 |
| <b>Chapter 3: Data Catalogs and Metadata Management</b> ..... | 12 | Getting People Involved .....                                      | 19 |
| Single Source for Shared Metadata .....                       | 13 | Understanding the Adoption Challenges.....                         | 19 |
|   |    | <b>Closing Thoughts</b> .....                                      | 21 |
|   |    | <b>About Eckerson Group</b> .....                                  | 22 |

# Introduction

The difficulties of data management have intensified at a steady pace over the past several years. The management complexities of big data, cloud hosting, self-service analytics, and data science can't be ignored. Effective data management has become a top priority for most organizations, but getting there is challenging. A data catalog has an essential role in overcoming these challenges.

Data catalogs were introduced to help data analysts find and understand data. Before data catalogs, most data analysts worked blind —

without visibility into existing data sets, their contents, or their quality and usefulness. As a result, analysts spent much of their time finding data, understanding data, and recreating data sets that already existed. Data catalogs were designed to address these issues.

From modest beginnings as a means to manage data inventory and expose data sets to analysts, the data catalog has grown in functionality, popularity, and importance. Modern data catalogs still meet the needs of data analysts, but have expanded their reach. They are now central to data stewardship,

data curation, and data governance. Data catalogs touch nearly everyone who works with data.

Success with data cataloging begins with fundamental knowledge of data catalog basics. You'll need to understand the what and why of data cataloging, the role and purpose of data curation, how data catalogs are a game-changer for metadata management, and the importance of collaboration and crowdsourcing. Ultimately, you'll need to plan for and drive data catalog adoption — getting all data stakeholders to participate in curation and cataloging processes and practices.

# Chapter 1: What is a Data Catalog?

## The What and Why of Data Cataloging

## Starting the Data Cataloging Journey

Data catalogs have quickly become a core component of modern data management. Organizations with successful data catalog implementations see remarkable changes in the speed and quality of data analysis, and in the engagement and enthusiasm of people who need to perform data analysis. By contrast, organizations without a data catalog often have these questions: What is a data catalog? Why do we need a data catalog? What does a data catalog do? These are all good questions and a logical place to start your data cataloging journey.

## Data Catalog Defined

A Data Catalog is a collection of metadata, combined with data management and search tools, that helps analysts and other data users to find the data that they need, serves as an inventory of available data, and provides information to evaluate the fitness of data for intended uses.

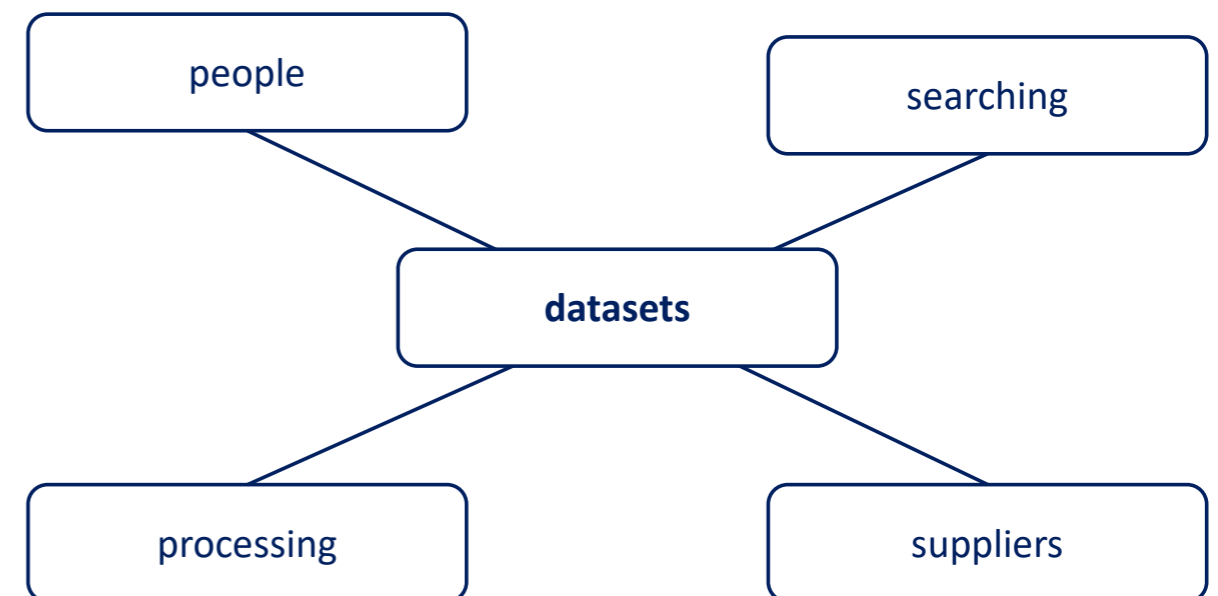
This brief definition makes several points about data catalogs — data management, searching, data inventory, and data evaluation — but all depend on the central capability to provide a collection of metadata.

Data catalogs have become the standard for metadata management in the age of big data and self-service analytics. The metadata that we need today is more expansive than metadata in the BI era. A data catalog focuses first on datasets (the inventory of available data) and connects those datasets with rich information to inform people who work with data. Figure 1 illustrates the typical metadata subjects contained in a data catalog.

Datasets are the files and tables that data workers need to find and access. They may reside in a data lake, warehouse, master data repository, or any other shared data resource. People metadata describes those who work with data —

consumers, curators, stewards, subject matter experts, etc. Search metadata supports tagging and keywords to help people to find data. Processing metadata describes transformations and derivations that are applied as data is managed through its lifecycle. Supplier metadata is especially important for data acquired from external sources, informing about sources and subscription or licensing constraints. We'll look more closely at catalog metadata in *Chapter 3: Data Catalogs and Metadata Management*.

**FIGURE 1 DATA CATALOG METADATA SUBJECTS**

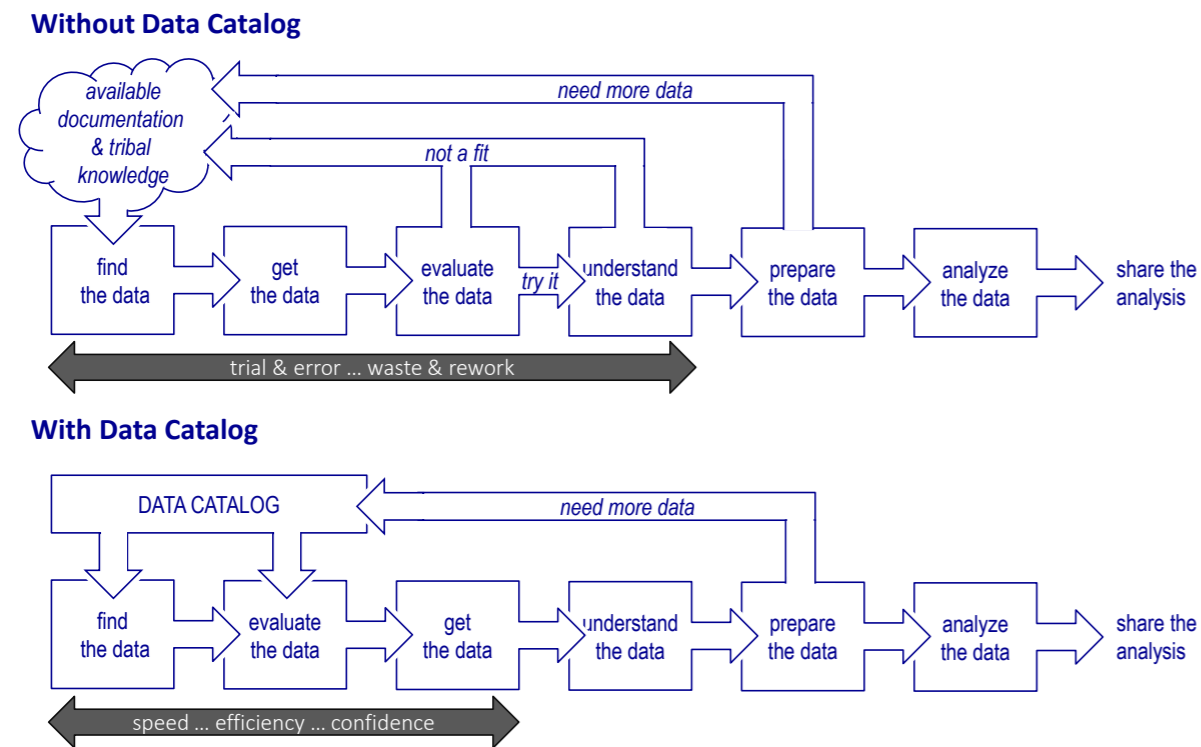


## Why Do We Need a Data Catalog?

The data management benefits of a data catalog become apparent by reflecting on the value of metadata and the capabilities that are created with comprehensive metadata. The greatest value, however, is often seen in the impact on analysis activities. We work in an age of self-service analytics. IT organizations can't provide all of the data needed by the ever-increasing numbers of people who analyze data. But today's business and data analysts are often working blind, without visibility into the datasets that exist, the

contents of those datasets, and the quality and usefulness of each. They spend too much time finding and understanding data, often recreating datasets that already exist. They frequently work with inadequate datasets resulting in inadequate and incorrect analysis. Figure 2 illustrates how analysis processes change when analysts work with a data catalog.

**FIGURE 2. ANALYSIS WITHOUT AND WITH A DATA CATALOG**



Without a catalog, analysts look for data by sorting through documentation, talking to colleagues, relying on tribal knowledge, or simply working with familiar datasets because they know about them. The process is fraught with trial and error, waste and rework, and repeated dataset searching that often leads to working with “close enough” data as time is running out. With a data catalog the analyst is able to search and find data quickly, see all of the available datasets, evaluate and make informed choices for which data to use,

and perform data preparation and analysis efficiently and with confidence. It is common to shift from 80% of time spend finding data and only 20% on analysis to 20% finding and preparing data with 80% for analysis. Quality of analysis is substantially improved and organizational analysis capacity increases without adding more analysts.

## What Does a Data Catalog Do?

A data catalog includes many features and functions that all depend on the core capability of cataloging data — collecting the metadata that identifies and describes the inventory of shareable data. It is impractical to attempt cataloging as a manual effort. Automated discovery of datasets, both for initial catalog build and ongoing discovery of new datasets is essential. Use of AI and machine learning for metadata collection, semantic inference, and tagging, is important to get maximum value from automation and minimize manual effort.

With robust metadata as the core of the data catalog, many other features and functions are supported, the most essential including:

- **Dataset Searching** — Robust search capabilities include search by facets, keywords, and business terms. Natural language search capabilities are especially valuable for non-technical users. Ranking of search results by relevance and by frequency of use are particularly useful and beneficial features.
- **Dataset Evaluation** — Choosing the right datasets depends on ability to evaluate their suitability for an analysis use case without needing to download or acquire data first. Important evaluation features include capabilities to preview a dataset, view data profiles, see user ratings, read user reviews and curator annotations, and view data quality information.

- Data Access — The path from search to evaluation and then to data access should be a seamless user experience with the catalog knowing access protocols and providing access directly or interoperating with access technologies. Data access functions include access protections for security, privacy, and compliance sensitive data.

A robust data catalog provides many other capabilities including support for data curation and collaborative data management, data usage tracking, intelligent dataset recommendations, and a variety of data governance features.

# Chapter 2: What is Data Curation?

## Managed Data Sharing



## More Than Shared Databases

Data curation is a term that has recently become a common part of data management vocabulary. Data curation is important in today's world of data sharing and self-service analytics, but I think it is a frequently misused term. When speaking and consulting I often hear people refer to data in their data lakes and data warehouses as curated data, believing that it is curated because it is stored as shareable data. Curating data involves much more than storing data in a shared database.

## What is Curation?

Let's set data aside for a moment and consider the meaning and the activities of curating. The word "curated" is used frequently today. The traditional use of the word is associated with collections of artifacts in a museum and works of art in a gallery. More recently we've started to use the term to describe managed collections of many kinds such as curated content at a website, curated music and videos available through streaming services, and curated apps through download services. Wired.com has described Apple's App Store as "curated computing."

Curation is the work of organizing and managing a collection of things to meet the needs and interests of a specific group of people. Collecting things is only the beginning. Organizing and managing are the critical elements of curation — making things easy to find, understand, and access.

## What is Data Curation?

If "curated" describes collections of things that are selected and managed to meet the needs of a specific group, then "curated data" is a collection of

datasets that is selected and managed to meet the needs and interests of a specific group of people. Note that the focus here is datasets – files, tables, etc. – that can be accessed and analyzed. The distinction between "collections of data" and "collections of datasets" is subtle but significant.

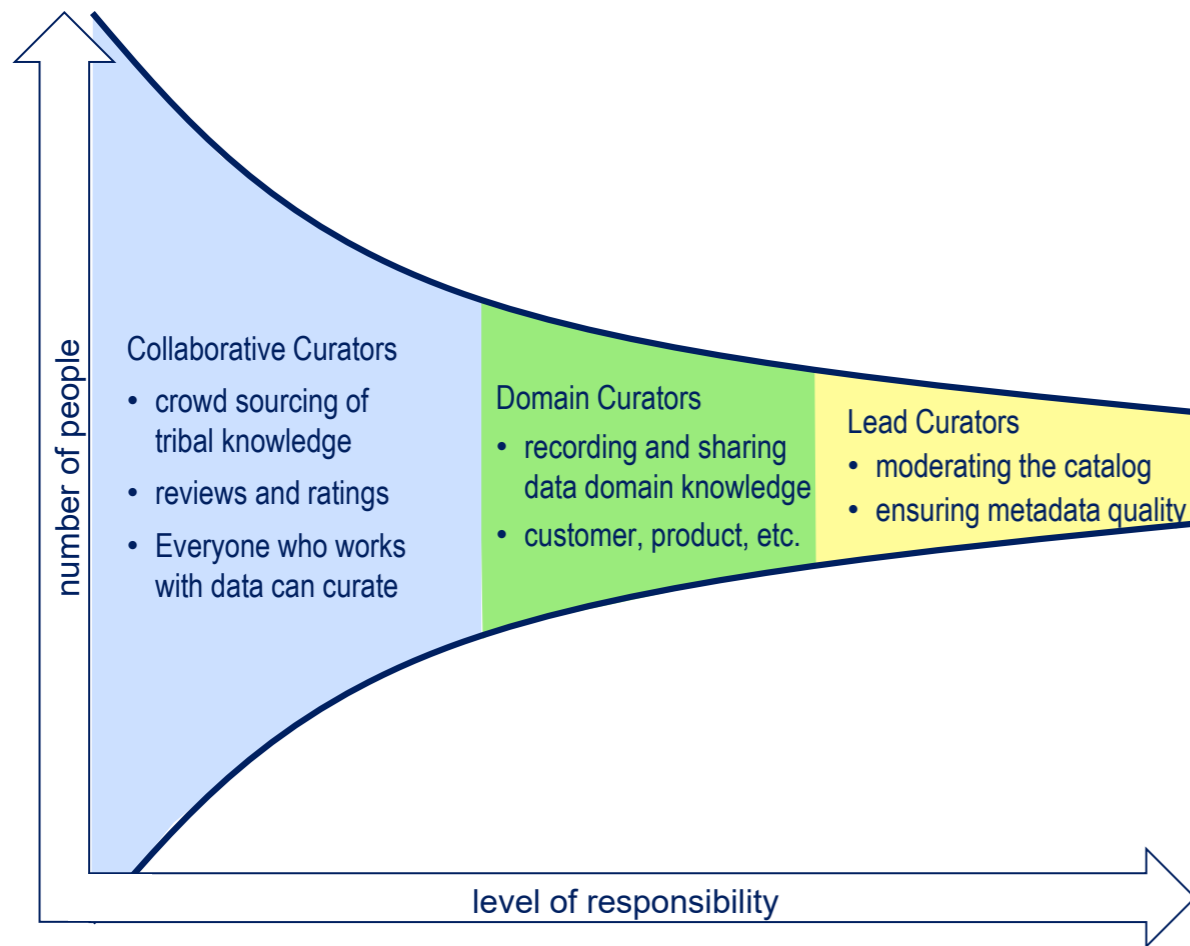
Data curation, then, is the work of organizing and managing a collection of datasets to meet the needs and interests of a specific groups of people. Collecting datasets is only the beginning. That is what we do when we store data in data warehouses or data lakes. But organizing and managing are the essence of data curation. Making datasets easy to find, understand, and access is the purpose of data curation — a purpose that demands well-described datasets. Data curation is a metadata management activity and data catalogs are essential data curation technology.

## Who Are the Data Curators?

A typical organization has many people doing data curation work (see figure 3) with varying degrees of responsibility and time commitment. Everyone who works with data has the opportunity to curate by sharing their knowledge and experiences. Crowdsourcing of tribal knowledge is an important part of curation practice. Collaborative data management is a necessity in the self-service world and knowledge sharing is the first step in creating collaborative culture. Curation collaborators will be large in number with a modest level of responsibility and time commitment.

Domain curators have subject expertise in specific data domains such as customer, product, finance, etc. Domain curators record and share data domain knowledge that helps data analysts to understand the nature of the data that they work with. The number of domain curators is substantially smaller than the number of collaborative curators, with greater level of responsibility and time commitment.

**FIGURE 3: CURATORS THROUGHOUT THE ORGANIZATION**



Most organizations will have one or very few lead curators who are responsible for moderating data catalog contents much as wiki moderators manage content. Lead curators have a high level of responsibility for metadata and catalog quality – responsibilities that require substantial time commitment.

## What About Data Stewards?

I frequently am asked about the differences between data curators and data stewards: Are they two names for the same role? Can data stewards be your data curators? Why do we need both stewards and curators? These are good

questions that are important when considering how to fit data curation into your organization. It is practical for the same individual to have both curation and stewardship responsibilities, especially at the level of domain curators. It is important, however, to recognize curation and stewardship and distinctly different roles, each with unique perspective about managing data. Some of the key differences are shown in the table below.

|       | DATA STEWARD                     | DATA CURATOR                           |
|-------|----------------------------------|--|
| Focus | entities and attributes          | categories and analysis variables      |
|       | relationships                    | data collections                       |
|       | databases                        | datasets                               |
|       | data elements                    | data pipelines and lineage             |
| Goals | business data requirements       | analytic data requirements             |
|       | priorities and data roadmap      | finding data when needed               |
|       | data-to-business alignment       | data-to-value alignment                |
|       | shaping data management policies | tracking data usage practices          |
|       | data quality improvement         | data quality evaluation                |
|       | security and privacy monitoring  | security and privacy policy compliance |
|       | data names and definitions       | descriptions, profiles, annotations    |
|       | business and technical metadata  | search and select metadata             |
|       | data-to-system mapping           | data-to-user tracking                  |

The roles of data steward and data curator are related and somewhat overlapping. Stewards and curators working together is a combination that maximizes the value of data across all use cases from enterprise reporting to analytics and data science. Stewardship and curation are both metadata management activities and

---

data governance roles. Data curation and data cataloging are important elements of modern data governance. They are complementary disciplines that are both essential in the age of self-service analytics.

Ultimately, data curation is a metadata management activity, and data cataloging is metadata management technology. But both approach metadata

very differently from metadata management practices of the past. *Chapter 3: Data Catalogs and Metadata Management* looks at metadata management in greater depth.

# Chapter 3: Data Catalogs and Metadata Management

## Knowledge Sharing for Data Management

## Single Source for Shared Metadata

Recall that we previously defined a data catalog as “**a collection of metadata**, combined with data management and search tools, that helps analysts and other data users to find the data that they need, serves as an inventory of available data, and provides information to evaluate the fitness of data for intended uses.” Although accurate, this definition overlooks one very important point: The data catalog serves as a resource of shared metadata. Everyone who has knowledge about data can share it through the catalog, and anyone seeking knowledge about data can find it in the catalog.

From modest beginnings as a means to manage data inventory and expose data sets to analysts, the data catalog has grown in functionality, popularity, and importance. Modern data catalogs — originated to help data analysts find and evaluate data — continue to meet the needs of analysts, but they have expanded their reach. They are now central to data analysis, data stewardship, data curation, and data governance — all metadata dependent activities.

## A New Approach to Metadata Management

It seems that everyone wants data management but most want to avoid metadata management. The distaste for metadata management is an artifact of past metadata approaches with disparate metadata collected by a variety of tools using proprietary formats and without integration. Metadata management in the BI era was painful, but we can't avoid the reality that metadata is essential to data management. Just as you need data about finances for effective financial management, you need data about data (metadata) for effective data management. You can't manage data without metadata.

As data management becomes more complex with data lakes, big data, self-service analytics, and data science, the role of metadata changes and the importance of metadata increases exponentially. Metadata that is current, accurate, and readily accessible is an imperative. Metadata disparity is not workable and metadata management as an afterthought is hazardous. We must actively manage metadata, and a data catalog is the right tool for the job. The data catalog has become the new gold standard for metadata and a cornerstone of data curation.

## Metadata in the Age of Self-Service

The real value of metadata is found in the answers it can provide. People who depend on data have questions about trustworthiness, latency, lineage, sensitivity, preparation, and much more. Sometimes they want to find others who know or have worked with the data to get human perspective. And they need to know about access, privacy and security constraints, cost, etc. Robust metadata ranging from data set names and properties to usage, access, licensing, and subject experts is the key to answering the many questions that data users and data managers will ask.

In today's self-service world, metadata is essential for three distinct groups of data management stakeholders:

- **Data consumers** need metadata to help them find data for reporting, analysis, and data science work, and to evaluate that data to ensure that they work with the right datasets.
- **Data curators** need metadata to observe data usage, understand the needs and interests of data consumers, and effectively manage the collection of shared data.

- **Data governors** (owners and stewards) need metadata to identify and protect sensitive data, trace data lineage, and establish trust in data.

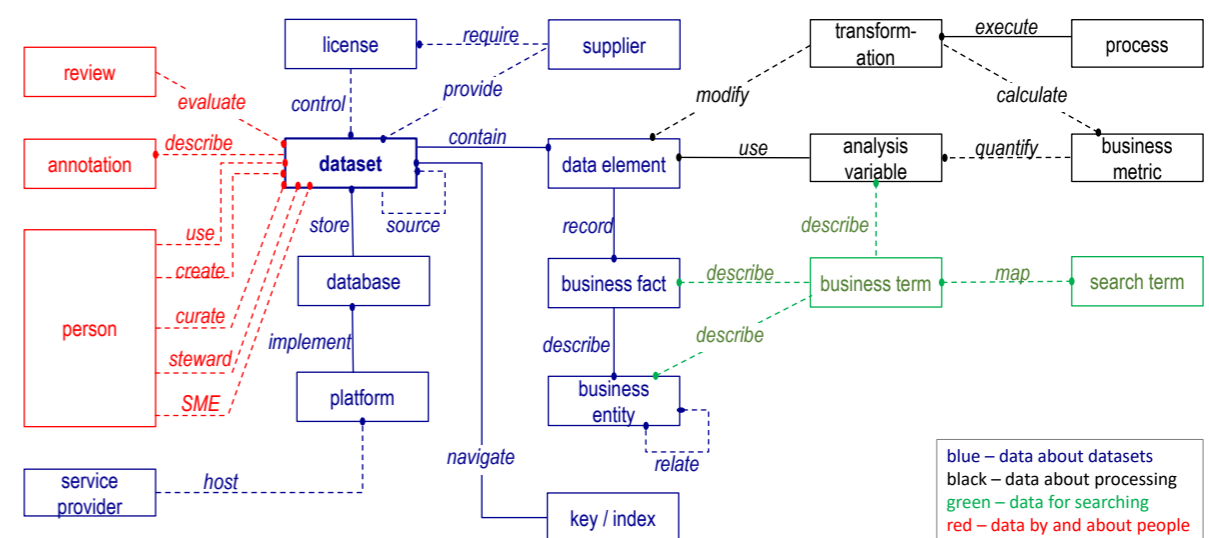
## Metadata and the Data Catalog

Metadata is the core of a data catalog. Every catalog collects data about the data inventory and also about processes, people, and platforms related to data. Metadata tools of the past collected business, process, and technical metadata, and data catalogs continue that practice. But data catalogs do much more. They collect metadata about datasets, metadata about processing, metadata for searching, and metadata for and about people. Figure 4 shows a logical data model that represents typical metadata content of a data catalog.

Data catalogs change the game and elevate best practices for metadata management with:

- **Crowdsourced metadata.** Much of catalog metadata is collected automatically by applying algorithms and machine learning. But sometimes the most valuable metadata is the knowledge and experiences of individuals and groups. Collecting that knowledge as user ratings, reviews, tips, and techniques enriches the metadata collection and converts tribal knowledge into a shared and enduring data management resource.
- **Data about people.** Data management and data analysis are ultimately human activities. Knowing which people have data roles and relationships and

**FIGURE 4. METADATA IN A CATALOG**



the nature of those roles is valuable. Data catalogs capture metadata to identify data users, data creators, data stewards, and data subject matter experts.

- **Automated metadata discovery.** Organizations with massive data holdings — literally tens of thousands of databases — simply don't know about all of the data they have. It is impossible to catalog a petabyte data estate without automated discovery.
- **Automated metadata discovery** is an important part of data cataloging. But much of the metadata in a data catalog is a result of crowdsourcing and collaboration. That is the subject of the next chapter.

# Chapter 4: Collaboration and Crowdsourcing

People and Culture in Data Cataloging



## Participative and Collaborative Data Management

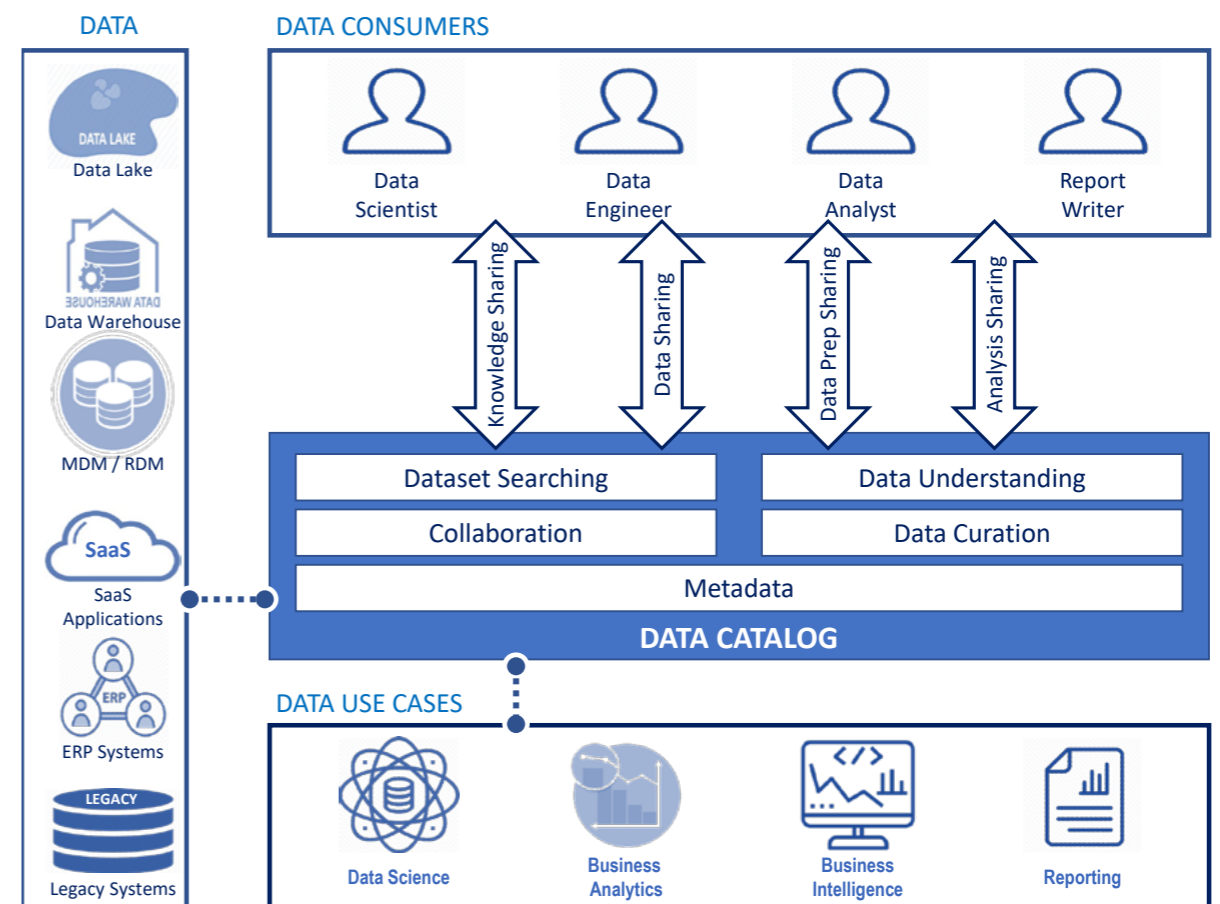
A core element of business today is the desire to become a data-driven organization. Most organizations aspire to that goal and many of them struggle. The key to data-driven success and maturity is data culture, and strong data culture begins with participation. Getting people at all levels from chief data officer to self-service data consumer to actively participate in data management activities is a barrier to building a strong and healthy data culture. A data catalog can be the catalyst that helps to break through the barrier with collaboration and crowdsourcing.

## Why Collaboration and Crowdsourcing — A Macro View

Collaboration is central to data-driven culture, creating an environment where no data stakeholders work in isolation, and where working together and sharing knowledge and experience is the norm. A robust and full-featured data catalog encourages collaboration and crowdsourcing with capabilities such as ratings, reviews, annotations, and deprecations. This is the human side of data cataloging that breaks down organizational silos and fosters a culture of sharing — knowledge sharing, data sharing, process sharing (data preparation), and analysis sharing. (See figure 5.) The data catalog becomes the centerpiece connecting people, data, and use cases in a way that improves both speed and quality of analysis.

Actively sharing knowledge, data, and experiences elevates data literacy and competencies of everyone involved. Working together exposes every individual to new information and different perspectives, often generating new ideas and sometimes sparking innovation.

FIGURE 5. SHARING VIA THE DATA CATALOG



## Why Collaboration and Crowdsourcing — An In-The-Trenches View

Everyone with a role in data management and everyone with data knowledge has opportunity and responsibility to collaborate in the processes and activities that make a data catalog valuable and informative. Data consumers, data curators, and data governors must all participate to create a culture of data sharing, metadata sharing, and knowledge sharing.

**Analysis and Reporting:** Finding the right data for a self-service reporting or analysis project is typically a difficult and time-consuming task filled



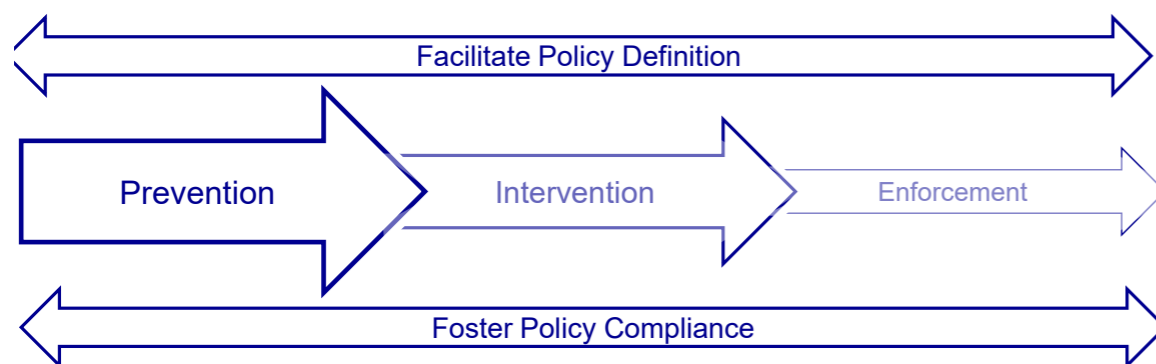
with unanswered questions. Users of data have questions about quality, trustworthiness, latency, lineage, and more. Sometimes they want to find others who know or have worked with the data to get a human perspective. Through collaboration the network of people willing to share their data knowledge rapidly expands. The effect is amplified with a data catalog that identifies data stewards, data coaches, data subject matter experts, and frequent users of datasets.

### FIGURE 6. RETHINKING DATA GOVERNANCE

#### Old Style Command-and-Control Data Governance



#### Modern Collaborative Data Governance



**Data Curation:** Data curation, as previously described, is the work of organizing and managing a collection of datasets to meet the needs and interests of

those who work with data. Recall the earlier description of three levels of data curators — lead, domain, and collaborative. Curators are the largest group, sharing and formalizing tribal knowledge and posting reviews and ratings to share their experiences when working with data. Crowdsourcing of tribal knowledge enriches catalog metadata and elevates the user experience for everyone who works with data. Crowdsourced knowledge from people who have worked with the data, consumer reviews, and usage tracking metadata help to evaluate and select the best-fit datasets for each unique analysis and reporting use case. Collaboration within and among the three levels of curators is an effective way to supercharge the richness and value of catalog metadata.

**Data Governance:** Adoption of self-service analytics has challenged conventional data governance practices. The top-down, command-and-control governance techniques of the past are at odds with the agility and autonomy interests of the self-service community. In the self-service world, collaborative data governance is an emerging and important practice. We must govern with the belief that most people want to do the right thing. The primary role of governance is to help them to know what is the right thing. Participation and collaboration are essential to fulfilling that role. (See figure 6.)

The data catalog is a core component of collaborative data governance. It provides a single point of reference for everyone who works with data. Everyone from chief data officers to self-service consumers sees the same metadata, and all have opportunity to share their knowledge, experiences, and perspectives about data. Crowdsourced, participative data governance is a natural fit for self-service organizations.

# Chapter 5: Driving Data Catalog Adoption

Engagement and Participation in Data Cataloging

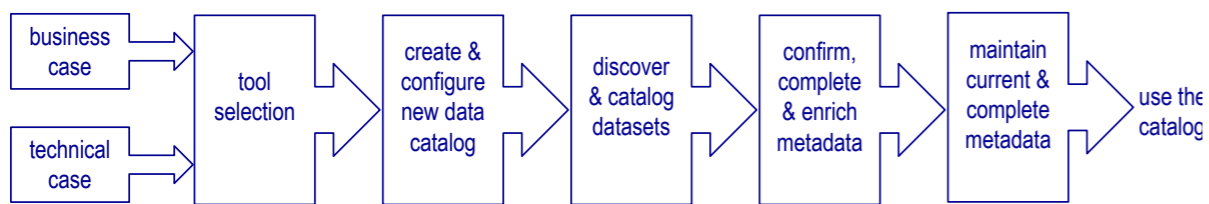
# Getting People Involved

Chapter 4: *Collaboration and Crowdsourcing* discussed the importance of participation by all data stakeholders as a key to getting maximum value from your data catalog. Many organizations, however, find data catalog adoption — getting people to participate — to be among the biggest challenges to data catalog success. Adoption is challenging, but understanding the causes of resistance and developing an adoption plan help to overcome those challenges.

# Understanding the Adoption Challenges

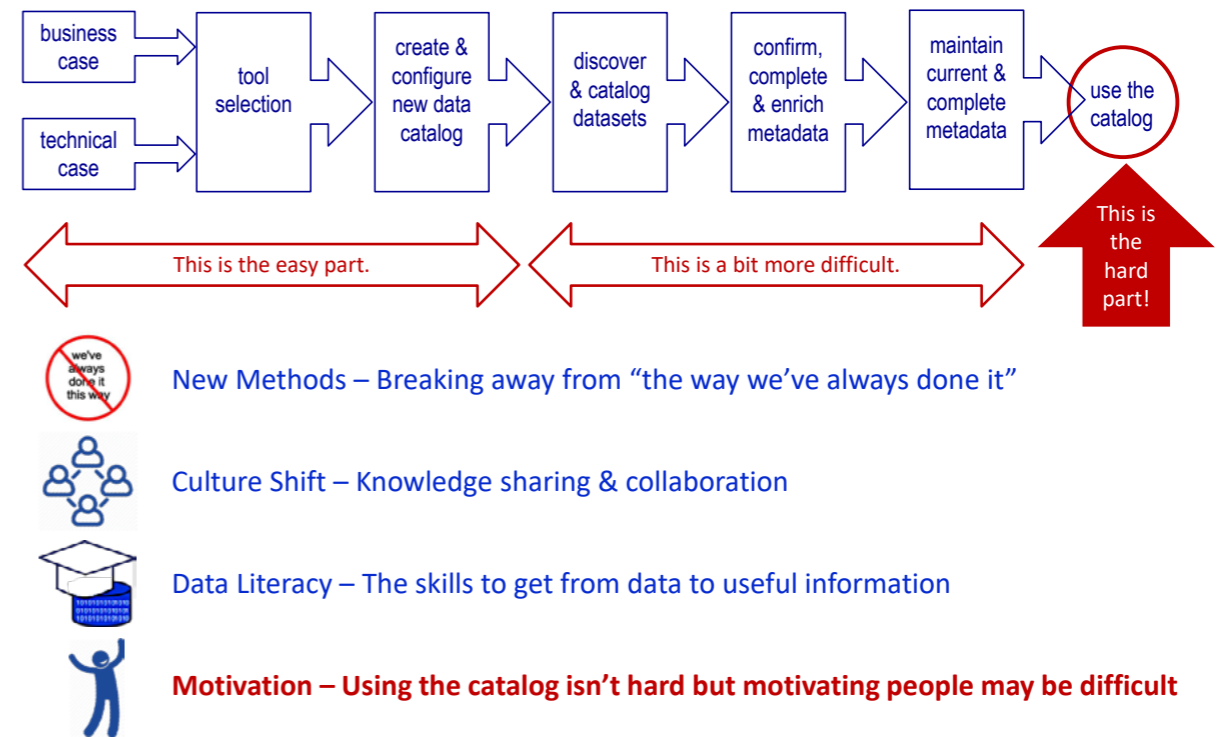
When planning for implementation, the human and cultural dimensions of data cataloging are often overlooked or subordinated to the process and technology dimensions. A typical data catalog implementation process begins by defining the business and technical case, proceeds through technology selection and installation, then moves on to data discovery and populating the metadata catalog. (See figure 7.) This build-it-and-they-will-come approach fails to engage people to actively use the catalog.

**FIGURE 7. DATA CATALOG IMPLEMENTATION**



The final step — use the catalog — often doesn't happen at the level expected for a variety of reasons. (See figure 8.) Predominant among those reasons:

**FIGURE 8. BARRIERS TO DATA CATALOG ADOPTION**



- **New Methods as a Barrier** — It is human nature to be anchored by “the way that we’ve always done it.” The shift to new ways of doing things pushes people away from the familiar and comfortable. Self-service data consumers may resist the data catalog and continue to rely on personal networks and tribal knowledge because it is what they know how to do. Using the data catalog requires them to learn new things, which can seem time-consuming and disruptive for busy people.
- **Culture Shift** — Data cataloging is most successful in a culture of data sharing, knowledge sharing, and collaboration. Behaviors such as “my data” mentality, territorialism, and knowledge hoarding are signs of an unhealthy culture that is a barrier to becoming a data-driven organization. A healthy data culture encourages collaboration and sharing, and discourages the unhealthy behaviors. Participation is a key element of data culture —

participation at all levels. Leadership visibly invests in data management and in growing data literacy throughout the organization. Staff are encouraged and incentivized to access and analyze data and to share their knowledge about working with data and share the insights that they derive from data.

- Data Literacy — Many line-of-business people have responsibilities that depend on data analysis but have not been trained to work with data. The skills to get from data to useful information — data selection, data understanding, data preparation, data analysis, data visualization, and data storytelling — are not native and natural for them. Their tendency is to do just enough data work to get by, and to do that work primarily in Excel

spreadsheets. The data catalog and access to abundant data often feels more like hazard than opportunity to these people.

- Motivation — Changing how you work and learning to use the data catalog can seem intimidating, time-consuming, or simply out of comfort zone. Most people will resist change until they see how it benefits them personally. What's-in-it-for-me (WIFM) is a typical response, especially when asked to do new things such as participate in metadata crowdsourcing and post ratings and reviews of datasets. WIFM is a major influence in resistance to data sharing, resistance to knowledge sharing, reluctance to participate in collaborative curation, and reluctance to post ratings and reviews.

## Closing Thoughts

Data catalogs are positioned to be an enduring part of the future of data management. They fill critical roles for data analysis, data curation, data governance, and data science. Effective use of a data catalog increases effectiveness and value derived from all of the other tools in your data and analytics technology stack. Data preparation, data analysis, and data science tools all see marked ROI increases when coupled with data cataloging. To realize the benefits of data cataloging, begin with the business and technical

case — know the what and why of data cataloging. Then put data curation practices into action to manage metadata, and encourage collaboration and crowdsourcing to enrich the metadata. Systematically and incrementally expand the reach of the data catalog, ultimately extending to all data consumers and stakeholders. With this approach to data cataloging you'll experience real business impact through increased capacity for data analysis, accelerated analysis, and improved quality and reliability of analysis results.



## About Eckerson Group

Wayne Eckerson, a globally known author, speaker, and advisor, formed [Eckerson Group](#) to provide data-driven leaders a cocoon of support during every step of their journey toward data and analytics excellence.

Today, Eckerson Group has three main divisions:

- **Eckerson Research** publishes insights so you and your team can stay abreast of the latest tools, techniques, and technologies in the field.
- **Eckerson Consulting** provides strategy, design, and implementation assistance to meet your organization's current and future needs.
- **Eckerson Education** keeps your data analytics team current on the latest developments in the field through three- and six-hour workshops and public seminars.

Unlike other firms, Eckerson Group focuses solely on data analytics. Our veteran practitioners each have more than 25 years of experience in the field. They specialize in every facet of data analytics—from data architecture and data governance to business intelligence and artificial intelligence. Their primary mission is to share their hard-won lessons with you.

Our clients say we are hard-working, insightful, and humble. We take the compliment! It all stems from our love of data and desire to serve—we see ourselves as a family of continuous learners, interpreting the world of data for you and others.

Accelerate your data journey. Put an expert on your side. Learn what Eckerson Group can do for you!



[Contact Us](#)

[Schedule a Call](#)



## About Alation

Alation, the data catalog company, is building a data-fluent world by changing the way people find, understand, trust, use and reuse data. The first to bring a data catalog to market, Alation combines machine learning and human collaboration to bring confidence to data-driven decisions. More than 150 organizations, including eBay, Exelon, Munich Re and Pfizer, leverage the Alation Data Catalog. Headquartered in Silicon Valley, Alation is funded by Costanoa Ventures, DCVC (Data Collective), Harmony Partners, Icon Ventures, Salesforce Ventures, and Sapphire Ventures.

For more information, [visit alation.com](https://alation.com).