



# Metadata Management Methodology

Learn why metadata management is foundational to building a data culture and five steps for starting a metadata management program.

## Why Manage Metadata?

Organizations have too much data. So much, in fact, that workers cannot be expected to find, understand, and use it effectively without the proper knowledge and tools.

Metadata is the foundation of that knowledge. It powers the tools that make data usable, and makes workers more effective at using data.

In a word, metadata delivers context. Metadata is information about data that can enable categorization, detail provenance, specify ownership, describe usage, and more. Examples of metadata include dates, sources, sizes, descriptions, and definitions. Amazon, the ubiquitous online retailer, guides shoppers with rich product metadata in the form of ratings, reviews, colors, sizes, weights, model numbers, and product details.

### Who Benefits from Metadata and How?

**Data consumers** need metadata to help them find data for reporting, analysis, and data science work 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 stewards** (owners and stewards) need metadata to identify and protect sensitive data, trace data lineage, and establish trust in data.

People who work closely with data — analysts, data scientists, IT teams, and others — rely on metadata for crucial context on how to use a given data asset. Digging deeper, those who generate, use, and are responsible for data all rely on metadata. Data owners and stewards use it to build trust in the data, data curators use it to understand the needs of data consumers, and consumers use it to find the right data for the right task.

But just as data has overloaded organizations, the accompanying metadata has grown to the point that it now requires dedicated attention to harness its power. For metadata to be useful, organizations must identify, capture,

and share it easily. In other words, metadata must be managed.

Metadata management is the process of capturing metadata from the underlying data, organizing it, and enabling organizations to manage, govern, and utilize the data more effectively. This requires a metadata management solution, typically a data catalog, for search, discovery, and governance to create an inventory of available data, provide information evaluating the fitness of that data, and facilitate metadata management.



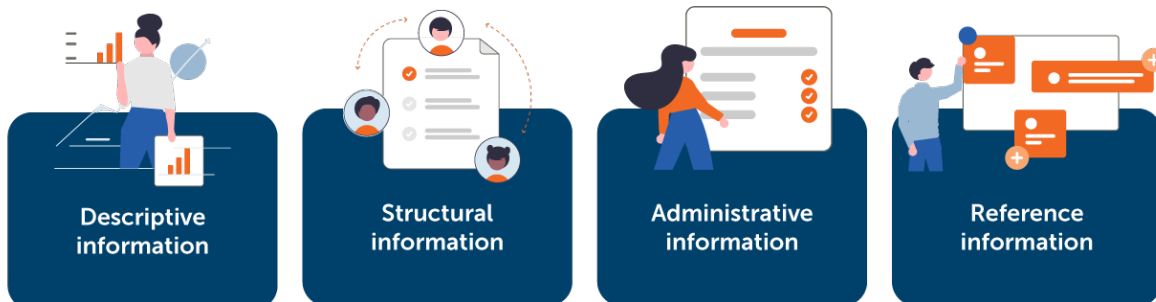
**Metadata management is becoming a critical functionality in practically all data-enabling technologies.**

**Gartner, Inc.**

**[Market Guide for Active Metadata Management](#)**

## Digging Deeper into Metadata

If metadata were simply categories and some descriptive information, it could be managed in a spreadsheet or lightweight database. But, especially for enterprise data and the opportunities and risks related to its use, the breadth of metadata requires a specialized tool. At the very highest levels, there are four key types of metadata:



1. **Descriptive** (or business) includes information on title, purpose, business intent, creation date, and creator. It is useful for cataloging and discovering data, and identifying which data is most appropriate for use.
2. **Structural** includes how data is formatted, with information on tables, pages, types, and relationships. It is useful for understanding how data is organized, determining if it can be combined with other data, and enabling relevant data discovery.
3. **Administrative** (or technical) includes access permissions, locations, data type, file name and size, and ownership information. It is useful for data governance, compliance, controls, and data management.
4. **Reference** includes information on quality, sources, processes used, schemas, and formulas. It is useful in determining how data can be utilized.

Descriptive metadata can be further defined as containing business-critical information such as definitions that establish a consistent shared meaning of data. This could include business and product names, descriptions of relationships between entities, rules that govern those relationships, and more. Those who capture and create these definitions are typically [data stewards](#) or someone who has formal accountability for data in the organization. Some examples include:

- **Classification data** covering security, privacy, product category, size, shape, color, or other pertinent details. Privacy information can include dispositions such as public, private, or restricted or regulatory classifications such as personally identifiable information (PII) and payments and financial data.

- **Data quality information** to gauge what can be trusted for different uses and why or why not. This could be related to accuracy, timeliness, source, or other details.
- **Popularity or behavioral information** to pull metadata from productivity tools such as Slack, Excel, Tableau, and others to reflect usage of data and level of adoption to increase usage, retire data that is not used, or flag data to better understand why it's not being used.
- **Lifecycle information** to ensure data is used effectively once generated and before it expires, with stages such as development, introduction, maturity, obsolescence, and retirement.
- **Classification data** covering security, privacy, product category, size, shape, color, or other pertinent details. Privacy information can include dispositions such as public, private, or restricted or regulatory classifications such as personally identifiable information (PII) and payments and financial data.

## 5 Steps To Get Started With Metadata Management

Extracting knowledge out of data is table stakes. Having a greater understanding of the data that is available and how it can be used increases organizational wisdom for the future as well as data capability velocity. Metadata is the enabler, but it needs to be accessible, searchable, and usable via metadata management. These five metadata management best practices can enable the effective, sustainable, and beneficial use of metadata across an organization.

### 1. Assign a metadata team

Success requires a dedicated team to develop a metadata management process and strategy, coordinate the rollout of processes and policies, and guide the selection of a metadata management tool. The metadata team should have experience in data management, data governance, and the organization's overall data landscape. They should also have the business acumen and relationships to connect a metadata management strategy with the organization's broader data and business strategies.

At this initial step it is helpful to define a scope for metadata management to specify the types and sources of data to be included, which business units or entities will be covered, typical data consumers, and more. This will then help to prioritize data assets and identify stakeholders. Also, to increase the value and impact of the program, set goals for metadata use beyond just search and discovery such as data governance, analytics, regulatory or other compliance, risk analysis, and quality.

### 2. Define a metadata strategy

Defining goals in step 1 will help determine desired outcomes. An organization may strive to launch a self-service analytics environment, instill a [data culture](#), enable more organizational agility, reduce request to their IT team, or something else entirely. The metadata strategy should take those goals into account to then begin to identify the required metadata, where it is located, and related technical requirements to access the metadata

and data. This is also a good time to determine where the metadata will be stored, how it will be accessed by metadata managers and data consumers, how it will be maintained, and who will be responsible.

As an example, the U.S. Department of Defense published its [DoD Data Strategy](#) to enhance “military effectiveness through access to accurate, timely, and secure data.” The document includes many references to metadata and how it supports the entity’s overall data strategy.

### 3. Adopt metadata standards and create a framework

Every organization should adopt a set of metadata standards to ensure uniformity and consistency as the basis for a metadata management process. Commonly accepted metadata standards do exist with a popular approach being the [Dublin Core Metadata Element Set](#) and the related [ISO 158369 standard](#).

The metadata standards will support the metadata management framework, which combines organizational structure and tools to establish a common taxonomy or language for everyone across the organization. Think of the framework as the organization’s data encyclopedia.

### 4. Deploy a metadata management tool

A purpose-built metadata management tool allows organizations to collect and utilize metadata. Such a tool typically takes the form of a data catalog as a repository and search tool. Modern data catalogs use AI and ML to automatically capture and categorize the metadata. Advanced metadata management tools also have capabilities to manage metadata management processes and administration, data policies, and data governance.

### 5. Scale metadata management

Rolling out a metadata management strategy across an organization can happen at once or in logical phases based on organizational departments, types of data, or other segmentations. It’s also important to continuously improve, adjust, and update the metadata management strategy, metadata management processes, and related policies, standards, and more.

## An Active Approach To Metadata Management

As the types and volumes of metadata grow, metadata management quickly becomes a domain unto itself, especially at larger organizations. To manage the sheer volume of metadata, the category of [active metadata](#) has emerged.

Active metadata signals a shift from manual to automated metadata management processes using artificial intelligence and machine learning (AI and ML). It enables continuous discovery and management of metadata across an organization's array of users, sources, and tools. The information is then analyzed to build recommendations based on the business outcomes achieved.

Learn more in [Gartner's Market Guide for Active Metadata Management](#). The analyst firm predicts that adopting active metadata tools and techniques will "decrease time to delivery of new data assets to users by as much as 70%."

## Tips for Creating a Metadata Management Framework

A [metadata management framework](#) defines how an organization uses metadata to enable more effective use of the underlying data. It is the basis for search and discovery that breaks down silos, makes data more accessible, and guides workers to the appropriate data. With data more accessible, data literacy increases because workers can better understand, interpret, and analyze data to improve decisions. For data governance, the framework sets clear expectations on how data is to be managed and used.

Developing a framework is critical to a successful metadata management program, but it can be accomplished quite easily. Key considerations include:

- **Keep it simple.** Begin at the highest levels to create a taxonomy that meets primary goals and limits the number of review cycles.
- **Assign clear roles and responsibilities to simplify data governance and convey accountability.** An administration team can also be assigned to develop the framework processes and strategy, coordinate the rollout, and guide the selection of a metadata management tool.
- **Leverage automation to streamline processes and reduce workload.** Data is constantly in flux, so automation can keep up with the tedious aspects of metadata management. Automation can also clean data, provide insights on data usage and expirations, track changes, categorize new data, and more.

## Determining Metadata Management Maturity

Whether developing a metadata management program from scratch or improving an existing program, it is helpful to benchmark a program's maturity against an accepted spectrum. Every initiative has room for improvement. Metadata management is also a relatively new discipline, so things like emerging data types, revised methodologies, and new technologies continue to shift expectations.

Higher levels of metadata management maturity lead to easier search and discovery and improved data literacy, all critical to building a data culture. Identifying and prioritizing data assets will define the schedule for expanding data management across the organization and ensuring visibility of the program's impact. But, recruiting and enabling data owners

to become stewards of the data is how a metadata management program will scale. Ensuring stewards own the responsibility and can reliably make needed updates will ease progression along the maturity spectrum.

A basic metadata management maturity model reflects the following progression organizations can expect to navigate:

| Maturity Level      | Characteristics                                                                                                                                                                                                                                                                                             | Recommendations                                                                                                                                                                                                                                                      |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Foundational</b> | Very little metadata is being curated. Data consumers rely on other team members to find data.                                                                                                                                                                                                              | Create a metadata administration team, create dashboards to measure metadata curation, and ensure data stewards are aware of their metadata management responsibilities.                                                                                             |
| <b>Developing</b>   | Metadata is documented only for key data sources. Data stewards understand the importance of metadata. Data consumers rely on ad hoc processes or other team members to find data.                                                                                                                          | Define a metadata management strategy, publicize related dashboards, and deploy a metadata management solution.                                                                                                                                                      |
| <b>Intermediate</b> | A metadata management solution is in place and metadata is documented for critical and other data sources. Data stewards actively participate in metadata management processes. Data consumers can easily find and use most data.                                                                           | Scale the metadata management program into more areas of the organization and to cover more data sources, monitor and report on progress, document major issues and develop remediation plans, and deploy advanced capabilities of the metadata management solution. |
| <b>Advanced</b>     | All critical and important data sources, both business and technical, are documented in the metadata management solution. Data stewards take ownership of metadata and governance efforts. Data consumers use efficient, structured, well-communicated routes to find most data without specialist support. | Utilize AI and ML capabilities to automate metadata management, create processes to monitor and improve metadata quality, and develop data awareness programs to increase data usage and inspire a data culture.                                                     |
| <b>Optimized</b>    | Data stewards document additional data sources and update existing metadata continuously in the metadata management solution, which is available across the organization. Data consumers can easily access, use, and reuse data without specialist support.                                                 | Continuously improve the metadata management program, incorporate more advanced features of the metadata management solution, and bring more data sources onto the solution.                                                                                         |

## Selecting a Metadata Management Solution

Metadata management typically relies on a data catalog as the software tool used to manage and organize metadata. Modern data catalogs go far beyond simple metadata management, search and discovery, governance, and collaboration. Expect top-tier solutions to learn from human behavior, auto-complete with search suggestions, warn of deprecated data as queries are created, house published queries for colleagues to reuse and adapt, and extend collaboration to third-party tools like Slack and Microsoft Teams. Also expect capabilities that empower programs to scale with analytics dashboards and reporting tools, social and gamification features to improve contributions, and connectivity across databases, data lakes, business intelligence tools, and cloud-based data repositories.

Other metadata management solution considerations include:

- **Ease of use** and intuitiveness to speed adoption and ensure usage across the organization.
- **Automation** to capture and make metadata highly usable such as automated metadata ingestion, execution of business rules, and even integration with spreadsheets to quickly capture existing manual metadata repositories.
- **Customization** and flexibility to meet unique needs without requiring changes to existing metadata management processes.

## Alation for Metadata Management

Alation provides a platform for metadata management best practices. By using its repository of metadata on information sources from across the enterprise, including datasets, business intelligence reports, visualizations, and conversations, the catalog helps workers quickly find and understand data to improve analytics, data governance, privacy, cloud transformation, and more.

**Get started today.** Alation dramatically improves the productivity of analysts, increases the accuracy of analytics, and enables confident data-driven decision making.

To learn more, [sign up for a weekly demo](#).