

# Internal Use Only

## Deploying Effective Metadata Management Solutions

**Published:** 2 April 2018   **ID:** G00347645

---

**Analyst(s):** Thornton Jared Craig

Metadata is a fount of untapped business value, but the flood of data impedes technical professionals from unlocking that value. Organizations that learn the art of using role-based frameworks to organize and process metadata can boost their data processing performance and business intelligence.

### Key Findings

- Emanating from multiple sources and operations in the data pipeline, metadata is increasing at an exponential rate. Even the best organizations struggle to effectively capture, centrally manage and derive value from all this metadata.
- Organizations that use role-based views can extract more relevant information from metadata than organizations that focus on unclassified sets of metadata generated by each source.
- Improved metadata usage enhances business, operational and governance functions by directly mapping the data life cycle to specific processes or data management initiatives. Hence, organizations that lack effective metadata management will miss these opportunities.

### Recommendations

Data and analytics technical professionals responsible for enterprise information architectures and systems should:

- Identify all data sources and processing engines for data movement and transformation as potential sources of metadata. A complete metadata view enables lineage and advanced data analysis.
- Categorize metadata based on roles. This approach enables metadata streams to map directly to specific data management use cases. Using a role-based model identifies where different types of metadata overlap and clarifies which data sources will add value for each use case.

# Internal Use Only

- Implement new metadata management solutions and harmonize them with existing metadata systems to centralize the view and management of all metadata. Use autodiscovery functions to crawl and capture metadata, enabling autogenerated views of data lineage.

## Table of Contents

Analysis.....	3
What Is Metadata Management, and Why Is It Important?.....	4
How Can a Role-Based Model Be Used to Identify and Categorize Metadata and Boost Its Value? .....	7
How Should Technical Professionals Address the Three Main Metadata Management Components — Capture, Categorize and Use?.....	9
Capturing Metadata.....	9
Categorizing Metadata.....	12
Using Metadata.....	15
What Business Value Can Be Derived From Metadata Features, Taxonomies and Frameworks?.....	17
Leveraging Metadata Features.....	17
Building Metadata Taxonomies and Semantic/Ontology Frameworks.....	20
What Kinds of Offerings in the Metadata Management Tool Market Can Help?.....	20
Recommendations.....	22
Conclusion.....	23
Gartner Recommended Reading.....	24

## List of Tables

Table 1. Categorizing and Mapping Metadata Types to Roles and Use Cases.....	14
Table 2. Useful Metadata Management System Features.....	18
Table 3. Selected Enterprise Metadata Management Tool Vendors and Offerings.....	21
Table 4. Selected Products With Metadata Management Features.....	22

## List of Figures

Figure 1. Types of Metadata.....	4
Figure 2. Metadata Management Answers Larger Questions About Data.....	6
Figure 3. Role-Based Model for Metadata Management.....	8
Figure 4. Capturing Metadata Within an Enterprise Metadata Management System.....	11

# Internal Use Only

Figure 5. Categorizing Metadata.....	13
Figure 6. Using the Metadata.....	16

## Analysis

The need to effectively manage metadata has become more acute than ever, as organizations become increasingly inundated with multiple sources and types of data. Each new data source and data operation creates a new set of metadata. Streaming ingestion, integration functions and data consumption all create metadata as data moves through this pipeline. More and more tools are producing and storing multiple sets of metadata — in both traditional and cloud data environments. Even logical data warehouses (LDWs) — designed to organize data — create more metadata, further adding to the proliferation of metadata across multiple repositories. Consequently, metadata often sits among disparate, uncoordinated taxonomies, hierarchies and datasets, and lacks standardization across tools and catalogs.

Considerable business value could be derived from this assorted metadata if it were managed and linked in the right way. However, such data is typically linked only to the lineage and provenance of individual data life cycles — not to roles or processes. Thus, many business benefits that could be gained often fail to materialize due to an inability to view and interpret this data in a useful way.

Given these challenges, implementing an effective enterprise metadata management solution is more important than ever. A well-implemented metadata management initiative can help organizations to:

- Gain a holistic view and better manage data throughout its full life cycle
- Enable data governance to support enterprise information management (EIM) initiatives
- Identify new data use cases and business opportunities
- Provide a central view of data lineage, transformation and movement
- Identify data quality issues
- Identify and resolve end-to-end performance and operational issues

Gartner recommends that organizations implement enterprise metadata management disciplines and solutions. In this report, we highlight key background information and guidance to help technical professionals launch an effective enterprise metadata management initiative within their organizations. Questions covered in this report include:

- What is metadata management, and why is it important?
- How can a role-based model be used to identify and categorize metadata and boost its value?
- How should technical professionals address the three main metadata management components — capture, categorize and use?

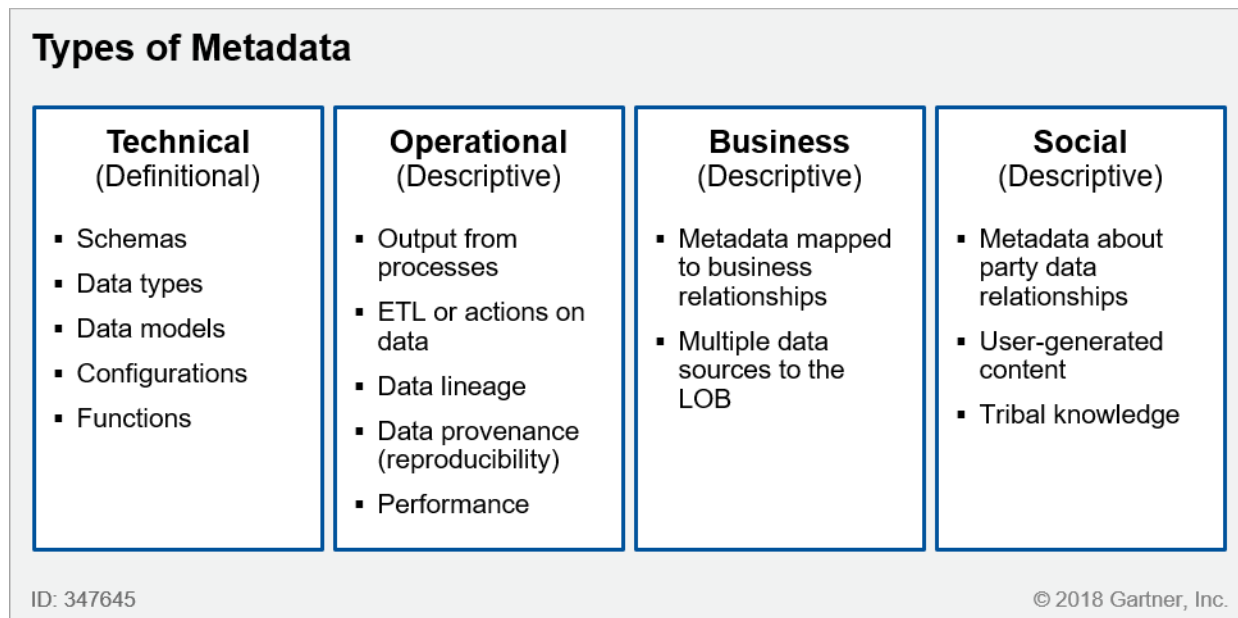
# Internal Use Only

- What business value can be derived from metadata features, taxonomies and frameworks?
- What kinds of offerings in the metadata management tool market can help?

## What Is Metadata Management, and Why Is It Important?

Metadata can be simply defined as "data about data." A more precise definition is that metadata describes various facets of an information asset in order to improve its usability throughout its life cycle. Figure 1 lists several categories and types of metadata.

Figure 1. Types of Metadata



ETL = extraction, transformation and loading; LOB = line of business

Source: Gartner (April 2018)

Metadata management is a process and discipline under which metadata is collected, governed, managed and organized — typically within an enterprise metadata management system. Within such a system, metadata is either imported or autodiscovered into a dedicated database. That database is linked to a hosted catalog, which shows types of available data and source descriptions.

Why is metadata management important? Because it can be used to derive additional, useful insights about data and relationships that weren't apparent before. Often, a metadata management system can correlate data from multiple data sources to create a much richer and more useful picture than what could be achieved from the data residing in the original context.

For example, in the case shown in Figure 2 below, information on users (TJ, Drue and Gene) appears in a variety of source data repositories. These include:

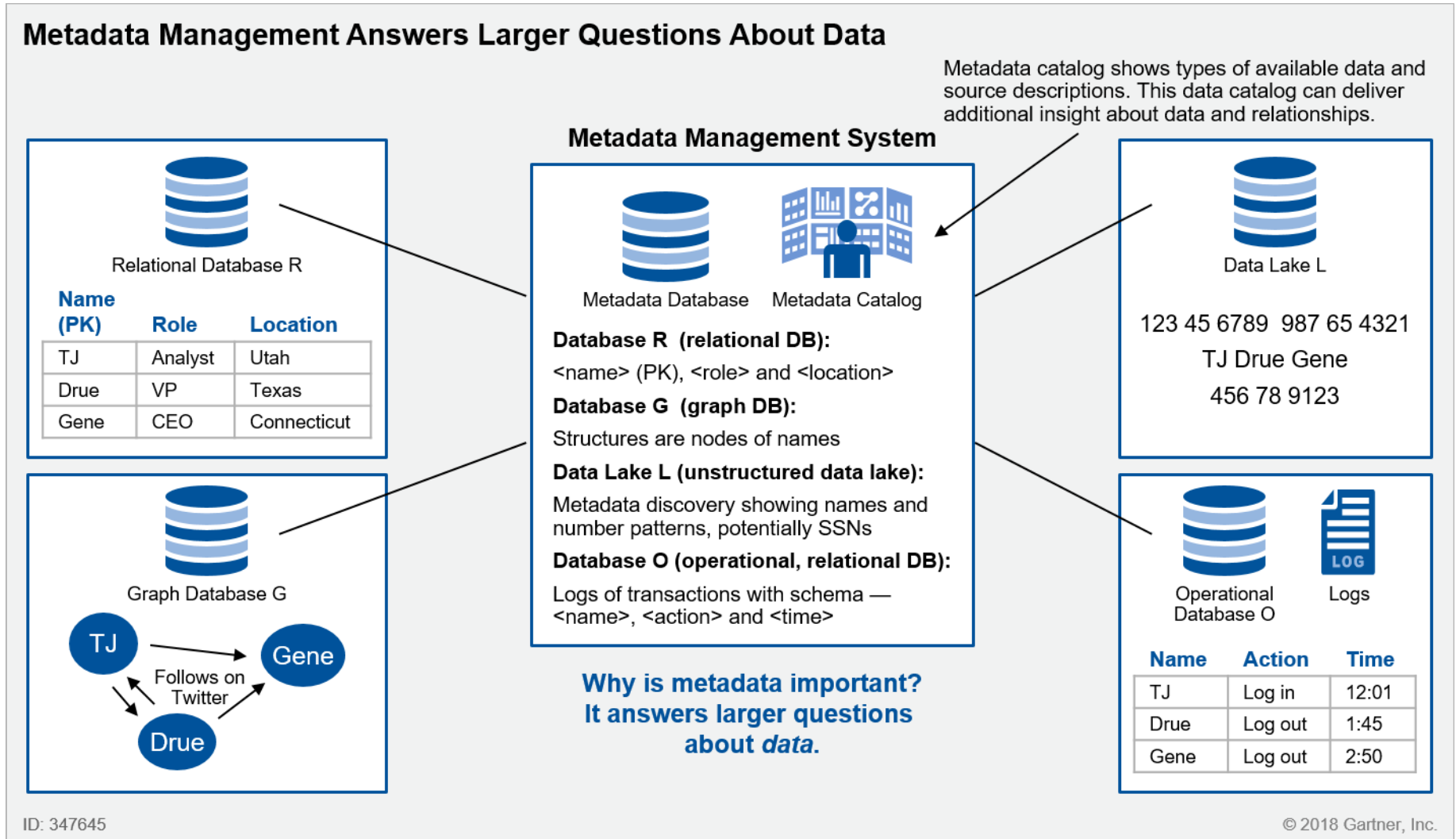
# Internal Use Only

- A relational database listing information like locations and roles
- An operational database logging system usage
- A graph database tracking social media relationships
- A data lake capturing other sundry information in an unstructured form

In this example, the ability to correlate information on these users across these data stores reveals an opportunity to map out their relationships. It also helps solve key questions by providing answers that couldn't be found by looking at individual data sources.

Figure 2 also reveals threats that can arise from not discovering and managing multiple, disparate sets of data. For example, the unstructured data in the lake appears to contain personally identifiable information (PII), possibly social security numbers. By linking this data to other data sources through matching or inferred external keys, metadata management exposes potentially sensitive data.

Figure 2. Metadata Management Answers Larger Questions About Data



DB = database; PK = primary key; SSNs = social security numbers

Source: Gartner (April 2018)

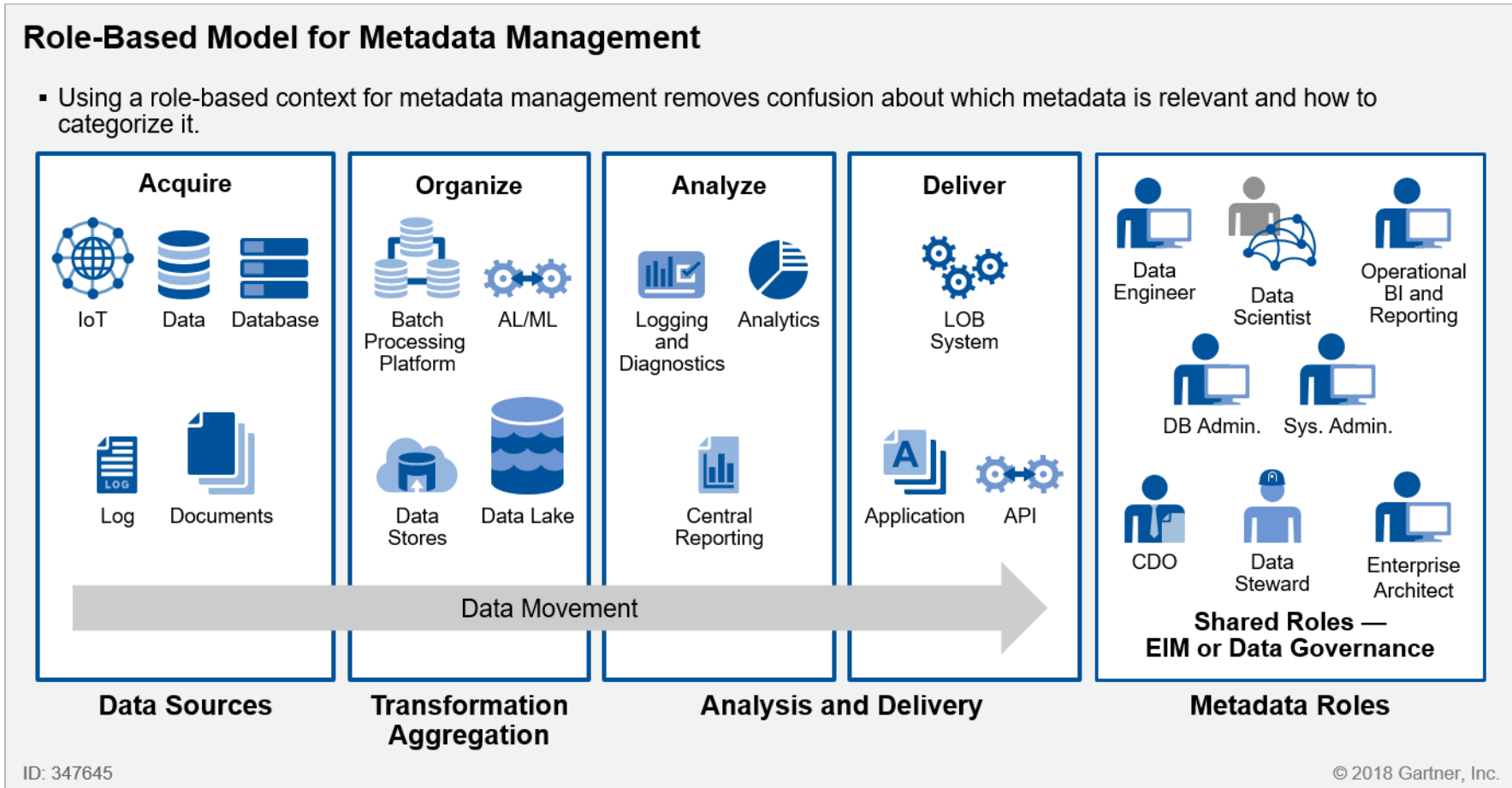
# Internal Use Only

## How Can a Role-Based Model Be Used to Identify and Categorize Metadata and Boost Its Value?

One key to turning the overload of enterprise metadata into useful business information is to map and categorize it in a way that will likely provide value to enterprise roles and processes. This is why Gartner recommends that organizations adopt a role-based model for identifying and categorizing relevant metadata as the core component of their enterprise metadata management initiative. The key benefit of using a role-based context for metadata management is that it removes confusion about which metadata is relevant and how to categorize it.

As shown in Figure 3, metadata is generated during multiple phases of the data life cycle (acquire, organize, analyze and deliver). It is also generated during the movement of data between phases. Different users may find only certain subsets of this metadata to be useful, and not others. For example, an artificial intelligence/machine learning (AI/ML) specialist may gain key insights from metadata generated during AI/ML processes carried out during the discovery, transformation and aggregation activities associated with the "organize" phase. A data scientist, on the other hand, may only be interested in metadata generated by one of the organization's data lakes during those activities. System administrators may only care about metadata associated with data performance or data logs that is generated during other phases in the life cycle. Meanwhile, more comprehensive collections of metadata may be valuable to roles that support broader, shared views of data — including roles responsible for EIM or data governance functions.

Figure 3. Role-Based Model for Metadata Management



BI = business intelligence; CDO = chief data officer; DB Admin. = database administrator; IoT = Internet of Things; Sys. Admin. = system administrator

Source: Gartner (April 2018)



# Internal Use Only

Many roles will consume overlapping sets of metadata. Roles that focus on the governance, lineage or provenance of data would be more interested in metadata related to broader views of data, encompassing its movement and management throughout the life cycle. For example, CDOs, data stewards or enterprise architects would likely be interested in a larger set of metadata that encompasses the majority of what's captured and analyzed in a metadata management system. Meanwhile, the metadata relevant to the roles with more limited interest may still overlap in some respects. For example, the pool of metadata linked to system administrators and data administrators would likely partially overlap, as would the metadata relevant to data scientists and data engineers.

This role-based model for metadata management is similar to traditional data security models, but requires shared access of overlapping sets of metadata. Sensitive data must always be protected, masked or anonymized, but roles may require access to complementary sets of metadata to discover useful relationships and lineage.

## How Should Technical Professionals Address the Three Main Metadata Management Components — Capture, Categorize and Use?

To make productive use of this role-based approach, technical professionals must implement an enterprise metadata management discipline, supported and built around the deployment of an enterprise metadata management technology solution. Three components are critical to such an initiative:

- Capturing metadata
- Categorizing metadata
- Using metadata

Each is examined separately in the sections below.

### Capturing Metadata

The first key goal is to discover metadata in the organization and to capture that data within an enterprise metadata management system. In a typical organization, multiple systems may already be generating metadata that isn't being captured. Such metadata-generating systems and sources may include LDWs, data lakes, databases, IoT systems, documents or system logs. In addition, organizations may have stand-alone tools that are designed to examine metadata, but only for small subsets associated with specific use cases. Thus, such tools cannot provide an enterprise-level view. Moreover, in-house-developed solutions, such as tools designed to automatically harvest log files or data movement information, may also generate smaller sets of metadata stored in stand-alone systems or repositories.

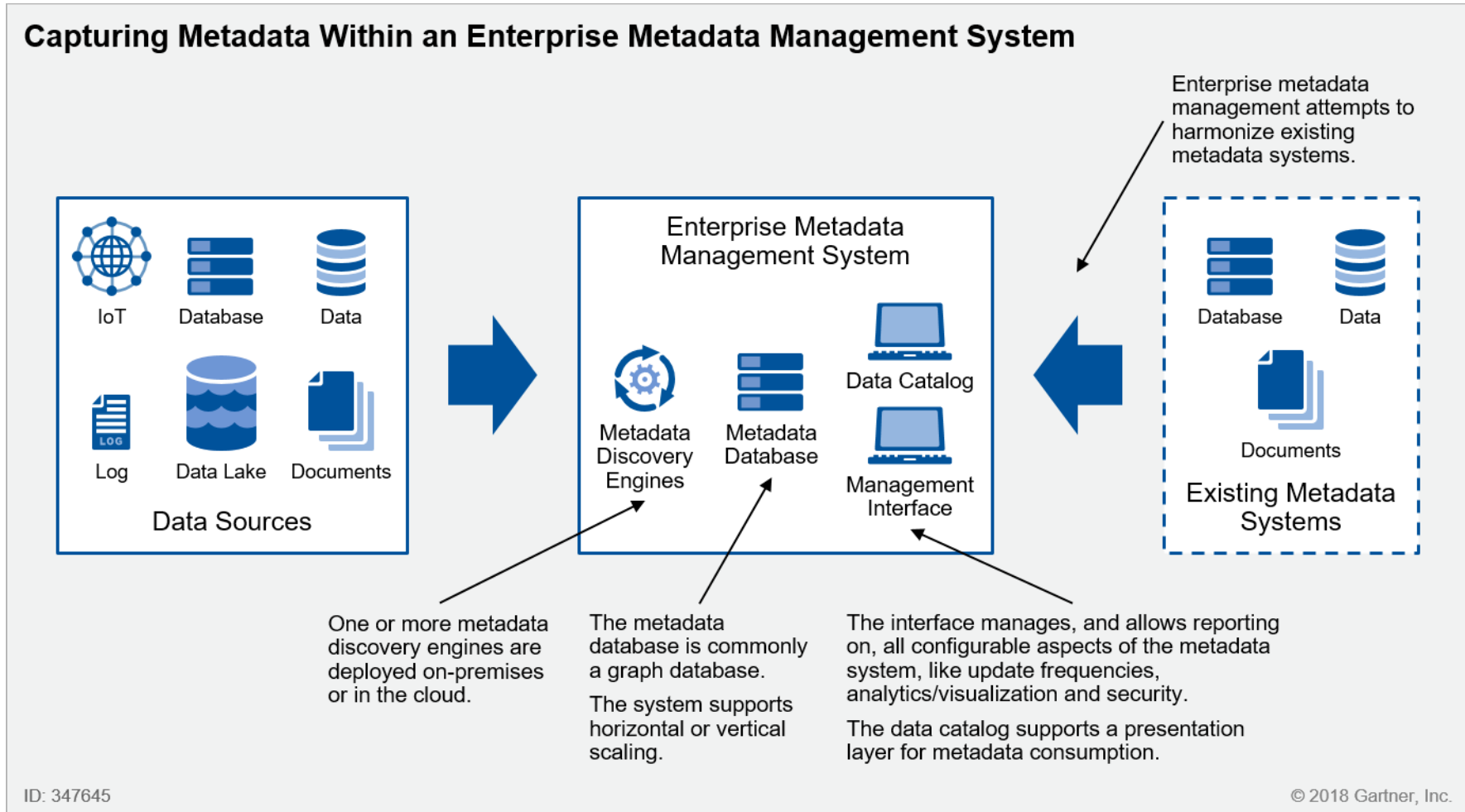
Offerings from metadata management tool vendors can be used to discover and capture the various types of metadata from these disparate sources. Vendors' enterprise metadata management solutions use intelligent engines that connect to source systems. These engines capture metadata

# Internal Use Only

tags, data definition language (DDL) and related data, and then create searchable metadata databases (see Figure 4).

*Autodiscovery* of metadata is one of the key benefits of modern metadata management systems. By using AI/ML algorithms to crawl existing systems, metadata tools can analyze complex environments in relatively short periods of time — especially when compared with traditional, human-based discovery and mapping of data sources, transformations and movements.

Figure 4. Capturing Metadata Within an Enterprise Metadata Management System



Source: Gartner (April 2018)

# Internal Use Only

Key components of an enterprise metadata management solution include:

- **Metadata discovery engines**, which can be deployed on-premises or in the cloud. These engines typically feature out-of-the-box connectors that will talk to most common systems containing metadata, and then discover and import that metadata into the system. Discovery engines use a variety of methods to capture metadata. These may include:
  - Reading a database management system (DBMS) catalog
  - Employing ML techniques that infer metadata by analyzing file contents and structures
  - Parsing ETL code
- **A metadata database**, which contains all of the aggregated metadata captured from the enterprise. This is typically a graph database that supports various types of horizontal or vertical scaling. These databases are often entirely cloud-based.
- **A hosted data catalog and a management interface**, which serve as the front end for the metadata database. These components offer a range of features to help users access and interpret metadata — including visualization and analytical tools that can produce maps of data lineage, transformations or movements in the enterprise.

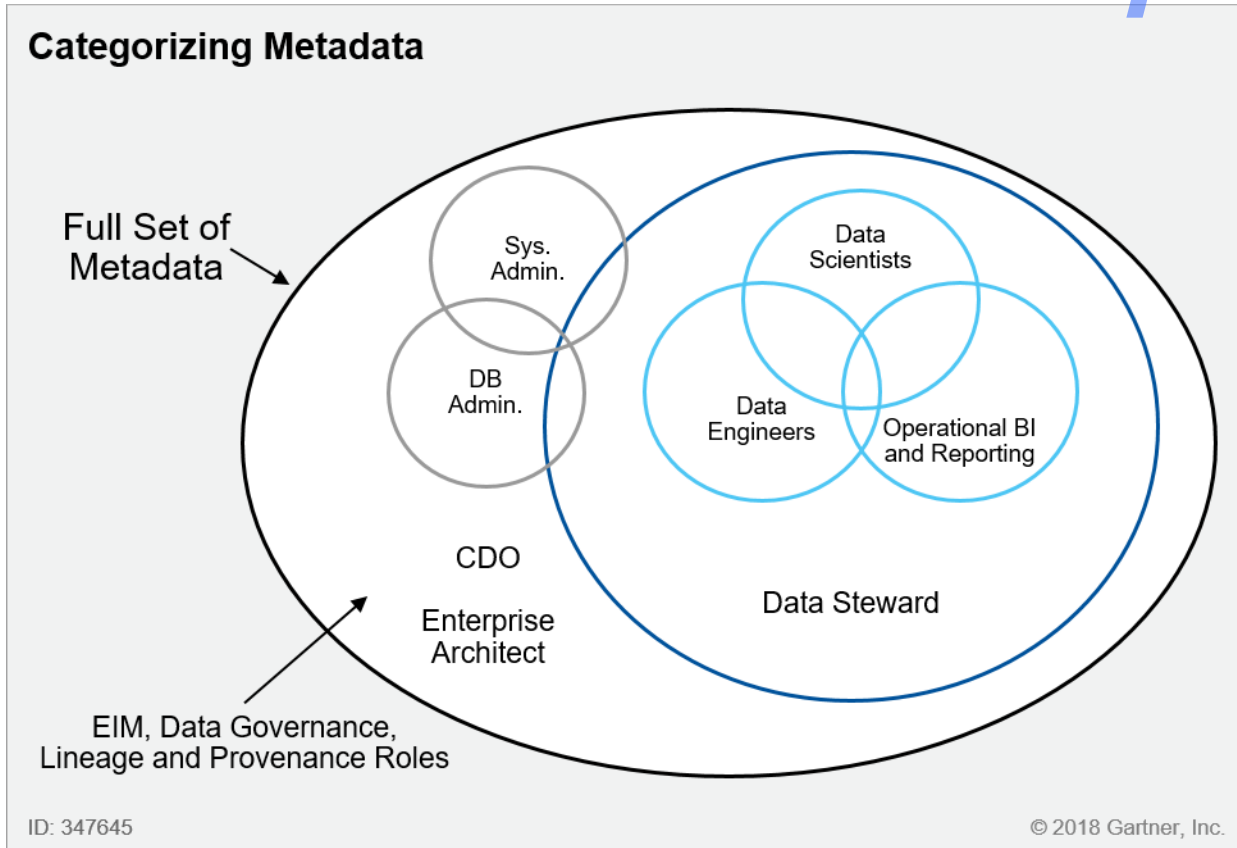
To meet the shared organizational requirements of EIM and data governance, Gartner recommends that enterprise metadata architectures standardize on a single metadata solution, aggregating metadata from individual use cases to build a global catalog. This will allow a holistic view of metadata to be captured, in preparation for later role-based categorization and productive use.

## Categorizing Metadata

The second key component of metadata management is categorization. The role-based method discussed earlier helps to sort metadata into useful categories that can help reveal insights and produce business benefits. Figure 5 and Table 1 show sample metadata categorizations based on the roles discussed earlier. Within such a categorization effort, different use cases should be identified for each role's productive utilization of metadata — along with the categories and specific types of metadata most relevant to each role. The key is to map the roles and use cases to appropriate metadata domains — both global and local.

# Internal Use Only

Figure 5. Categorizing Metadata



Sys. Admin = system administrator; DB Admin. = database administrator

Source: Gartner (April 2018)

# Internal Use Only

Table 1. Categorizing and Mapping Metadata Types to Roles and Use Cases

Roles	Use-Case Descriptions	Metadata Categories	Metadata Usage
<b>CDO and Enterprise Architect</b>	<ul style="list-style-type: none"> <li>Global metadata roles for data discovery for EIM and data governance</li> <li>Data discovery for specific or global domains</li> </ul>	<ul style="list-style-type: none"> <li>Technical</li> <li>Operational</li> <li>Business</li> <li>Social</li> </ul>	<ul style="list-style-type: none"> <li>Data design lineage</li> <li>Data provenance</li> <li>EIM</li> <li>Environment discovery</li> <li>Mapping business metadata</li> </ul>
<b>Data Engineers and Data Scientists</b>	<ul style="list-style-type: none"> <li>Unstructured or semistructured data discovery</li> <li>Big data processing</li> <li>AI/ML functions</li> <li>Feature engineering</li> <li>Data quality</li> </ul>	<ul style="list-style-type: none"> <li>Technical</li> <li>Business</li> <li>Social</li> </ul>	<ul style="list-style-type: none"> <li>Schema on read</li> <li>Data wrangling</li> <li>AI/ML design and implementation</li> </ul>
<b>Data Steward</b>	<ul style="list-style-type: none"> <li>Implementing and maintaining data governance</li> <li>Data quality initiatives</li> </ul>	<ul style="list-style-type: none"> <li>Technical</li> <li>Operational</li> <li>Business</li> <li>Social</li> </ul>	<ul style="list-style-type: none"> <li>Capturing and improving data quality</li> <li>Data real-time lineage</li> </ul>
<b>Database Administrator and System Administrator</b>	<ul style="list-style-type: none"> <li>Operational monitoring</li> <li>Performance metrics</li> </ul>	<ul style="list-style-type: none"> <li>Technical</li> <li>Operational</li> </ul>	<ul style="list-style-type: none"> <li>Performance</li> <li>Monitoring</li> </ul>
<b>Operational BI and Reporting</b>	<ul style="list-style-type: none"> <li>Specific business use cases</li> <li>LOB applications</li> <li>ETL, data cubes and dashboards</li> <li>Behavior and marketing analysis using social metadata</li> </ul>	<ul style="list-style-type: none"> <li>Business</li> <li>Operational</li> <li>Social</li> </ul>	<ul style="list-style-type: none"> <li>BI analysis</li> <li>Application development</li> <li>Business catalogs</li> <li>Business taxonomies</li> <li>Social media</li> </ul>

Source: Gartner (April 2018)

Multiple systems often produce disparate sets of metadata. As shown in Figure 5 and discussed earlier in this report, there may be some overlap in metadata domains. Enterprise metadata management systems attempt to harmonize all these sets of metadata into a holistic view. Unless an enterprise metadata management tool can aggregate the metadata domains, stand-alone

# Internal Use Only

systems still maintain sets of domain-specific metadata. Unresolved conflicts in metadata taxonomies often hinder integration of metadata domains. This is not ideal and does not support EIM functions like data governance.

To support EIM and data governance initiatives, organizations need to implement a complete view of metadata. As shown earlier in Figure 2, linking the disparate data sources enables organizations to discover potential security risks or data governance issues. In the Figure 2 example, the numbers in the data lake appear to be SSNs. Although these numbers are not tied to other data in the unstructured source, mapping the metadata between sources reveals potential security risks for that data.

Consider the following when categorizing metadata:

- **Category overlaps** may include all types of metadata (technical, operational and business), or a single subset. Two sets of technical metadata may not be in the same domain, but may share other types of metadata, such as operational.
- **Business metadata** is often the most valuable in linking data lineage back to functional processes, but is also difficult to achieve. When looking at data lineage from a low-level ETL or data movement perspective, always focus on the larger business questions being addressed by the overall data flow mapping.
- **Metadata domain conflicts** occur between separate taxonomies in stand-alone data sources or metadata systems. An enterprise metadata management system can create a shared taxonomy between these sources. Resolving these conflicts is challenging, due to lack of consensus. As a result, an enterprise metadata management solution may struggle to capture a holistic view of metadata, but can still lay the groundwork for a centralized metadata management system.
- **Social metadata** is a fourth category of metadata that includes user-generated content from multiple social platforms. This type of metadata is viewed as business-use-case-specific, but many organizations leverage social metadata to capture behavioral information about customers and employees.

## Using Metadata

Once all metadata is captured and categorized within the solution, it can be put to productive use by:

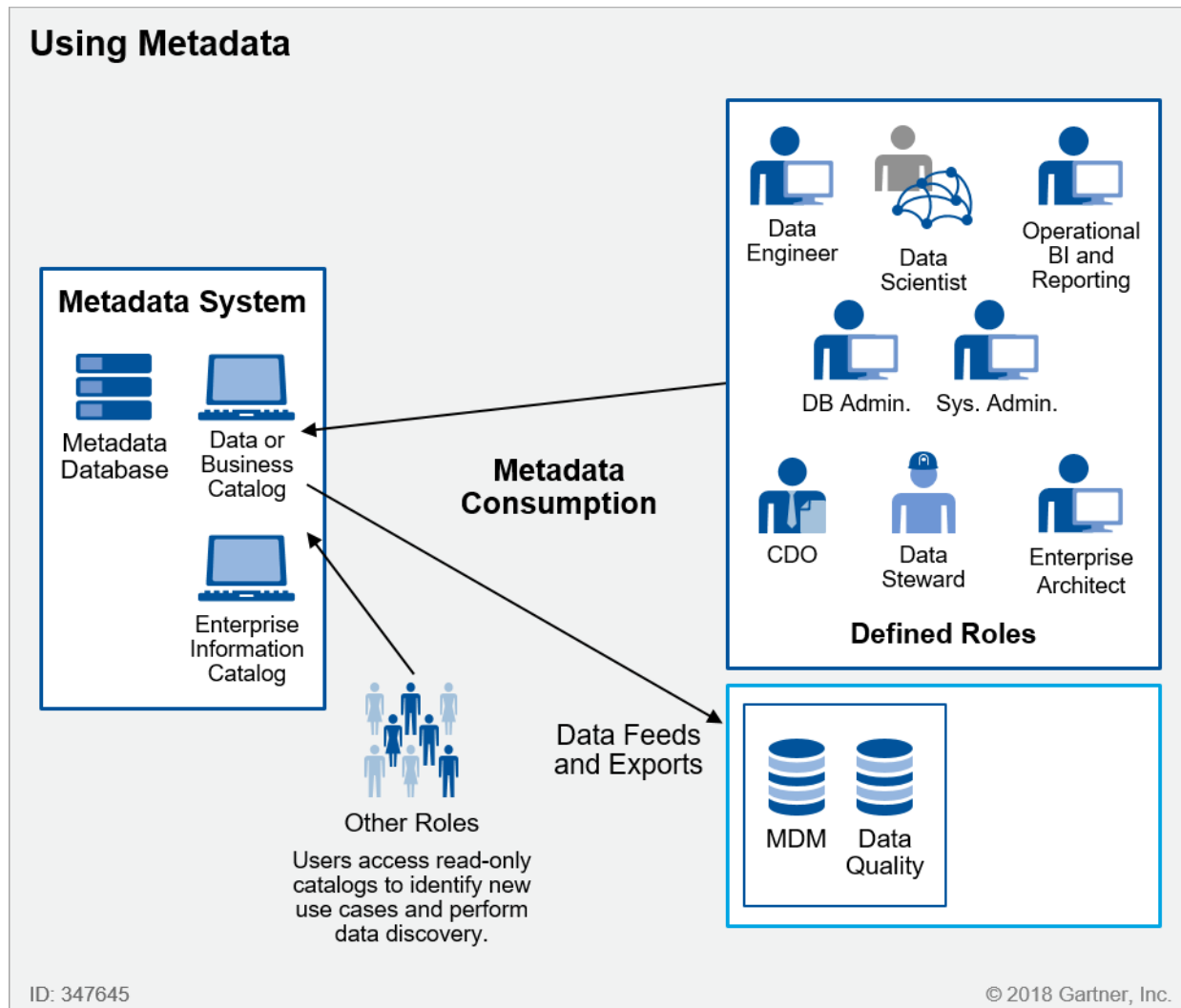
- Linking metadata to specific business roles and operational processes
- Using holistic metadata lineage to enable EIM, master data management (MDM) and data governance functions

Figure 6 shows how metadata can be put to practical use via:

# Internal Use Only

- **"Pull" access** from various user roles. Any of the defined roles can connect to a business catalog or enterprise information catalog to peruse the metadata. Access management controls ensure that specific sets of metadata are limited to specific users.
- **"Push" dissemination** in the form of data feeds and exports to EIM, data governance and data quality systems.

Figure 6. Using the Metadata



Source: Gartner (April 2018)

In addition, once you have a browsable, read-only metadata catalog, you can open it up to other people in the environment (assuming there are no PII or compliance restrictions). These people may, in turn, discover use cases, mappings or relationships in the data that nobody knew about previously. These users may include not only data scientists, but also other business users who may seek to gain insights via such analysis and discovery.



# Internal Use Only

In this way, the enterprise metadata management initiative can provide the insight needed to gain many of the benefits cited earlier, such as:

- Discovering new data use cases and business opportunities
- Analyzing data to highlight data quality issues in the pipeline
- Using metadata-enabled data lineage and prototyping capabilities to identify and resolve end-to-end performance and operational issues

## What Business Value Can Be Derived From Metadata Features, Taxonomies and Frameworks?

---

Once you have a metadata management architecture in place and an enterprise metadata management solution up and running, you can leverage several useful features and functions to gain additional business benefits.

### Leveraging Metadata Features

Many of the features available in today's enterprise metadata management tools provide ways to gain better insight into your enterprise data and apply it more usefully in the organization. Autogenerated tagging capabilities, for example, can provide value through added categorization. In addition, the lineage of data can be first designed, and then tracked in real time. Once these data movements are observed and analyzed, further analysis can be performed to:

- Determine the impact of different data movements and interactions
- Prototype various scenarios under which these impacts may play out

Another notable feature is the data "confidence" level, which is a measure of the data's quality. This feature seeks to determine how valid or fit for purpose the data is, based on how metadata definitions of that data compare with the reality reflected in actual data samples.

Table 2 lists descriptions and examples of these and other metadata management features that may yield additional business value.

# Internal Use Only

Table 2. Useful Metadata Management System Features

Feature	Description	Examples
<b>Age</b>	<ul style="list-style-type: none"> <li>Tracking when the data was ingested, moved or updated</li> </ul>	<ul style="list-style-type: none"> <li>Updates to data</li> <li>&lt;ingestion date&gt;</li> <li>&lt;copied&gt; and &lt;target&gt;</li> <li>&lt;last modified&gt;</li> <li>&lt;TTL&gt; (in data lakes, data assets may have a fixed time to live [TTL] or life cycle)</li> </ul>
<b>Assets</b>	<ul style="list-style-type: none"> <li>Data on metadata assets</li> <li>What the asset is, and when it was created</li> <li>Other unique identifiers</li> </ul>	<ul style="list-style-type: none"> <li>&lt;name&gt;</li> <li>&lt;date created&gt;</li> <li>&lt;org&gt;</li> <li>&lt;location&gt;</li> <li>&lt;description&gt;</li> </ul>
<b>Confidence</b>	<ul style="list-style-type: none"> <li>Calculation of data validity or quality</li> <li>Based on factors of defined usage, and derived from sample data</li> <li>For example, five valid samples out of 10 total equals a confidence level of 50%</li> </ul>	<ul style="list-style-type: none"> <li>Generated percentage values</li> <li>Confidence equals 0%, 25%, 50%, 100%, etc.</li> </ul>
<b>Data Governance</b>	<ul style="list-style-type: none"> <li>Capturing sensitive data assets (HIPAA, PII, EU GDPR-compliant, etc.)</li> </ul>	<ul style="list-style-type: none"> <li>Data governance models</li> <li>EIM architectures</li> <li>&lt;sensitive&gt;</li> <li>&lt;masked&gt;</li> <li>Aggregation of lineage for a data asset to support governance</li> </ul>
<b>Impact Prototyping</b>	<ul style="list-style-type: none"> <li>Impact analysis of data interactions</li> <li>Prototyping of data movement scenarios</li> </ul>	<ul style="list-style-type: none"> <li>Generated model of data impacts</li> <li>Lineage map over time</li> </ul>
<b>Lineage</b>	<ul style="list-style-type: none"> <li><i>Design lineage</i> refers to mapping data flow in a design mode; this allows for creating scenarios of how movement occurs in the full life cycle of the data</li> <li><i>Real-time lineage</i> tracks history as data moves; this is used to capture</li> </ul>	<ul style="list-style-type: none"> <li>Lineage relates to data movement</li> <li>Design lineage is used in architecture</li> <li>Real-time lineage is used in operations, in troubleshooting or for feedback to a design model</li> </ul>

# Internal Use Only

Feature	Description	Examples
	real-time changes for historical reference	
<b>Permissions</b>	<ul style="list-style-type: none"> <li>Identifying who or what process has access to the data</li> </ul>	<ul style="list-style-type: none"> <li>Defined permissions models</li> <li>&lt;owner&gt;</li> <li>&lt;read&gt;, &lt;write&gt;, &lt;update&gt; and &lt;delete&gt;</li> </ul>
<b>Statistics</b>	<ul style="list-style-type: none"> <li>Standard statistical values, including outliers and data volume</li> </ul>	<ul style="list-style-type: none"> <li>Minimum and maximum</li> <li>Skew and length</li> <li>Identified outliers</li> <li>Data volume</li> </ul>
<b>Tags</b>	<ul style="list-style-type: none"> <li>Manual or generated tags identifying types of existing and custom metadata</li> </ul>	<ul style="list-style-type: none"> <li>&lt;operational&gt;</li> <li>&lt;technical&gt;</li> <li>Custom tags to support categories or operations               <ul style="list-style-type: none"> <li>&lt;objectID&gt;</li> <li>&lt;department&gt;</li> <li>&lt;refreshed on&gt;</li> </ul> </li> </ul>
<b>Usage</b>	<ul style="list-style-type: none"> <li>Tracking when the data was used and for what purpose</li> </ul>	<ul style="list-style-type: none"> <li>Data consumption</li> <li>DDL to describe audit logs               <ul style="list-style-type: none"> <li>&lt;created_on&gt;</li> <li>&lt;accessed_on&gt;</li> <li>&lt;accessed_by&gt;</li> </ul> </li> </ul>

GDPR = General Data Protection Regulation; HIPAA = Health Insurance Portability and Accountability Act

Source: Gartner (April 2018)

As shown in Table 2, tags are a simple and commonly used method to identify and categorize metadata. Implementations of a metadata management system may use unique tags (both internal and user-defined). Database structures may also map facets of the actual metadata relationships. For example, commonly used graph databases may track a data asset as a node, and use edge relationships to capture lineage.

# Internal Use Only

## Building Metadata Taxonomies and Semantic/Ontology Frameworks

Many metadata tools can discover your data, and then autobuild taxonomies using AI and ML. This is particularly true of the tools that came out of the unstructured/data lake arena. Internal semantic layers can be developed from the discovered sets of metadata. Rediscovery and rebuilding of the metadata database allow semantic-layer changes to be captured over time.

Vendor solutions use classification rules for AI/ML training to discover and update classifications. These can be seeded in advance with your organization's own classification rules (e.g., for finance). Alternatively, the classification algorithms and routines can be run independently to detect and infer relationships, and then build and present an original data taxonomy based on this analysis.

Such taxonomies are typically high-level ones (e.g., for business, HR or finance). Business metadata may require custom taxonomies to identify specific relationships or custom values for a data model. Many vendor solutions also offer out-of-the-box taxonomies geared toward specific markets or business functions.

In addition, many vendor solutions support the creation of semantic and ontology frameworks based on common data modeling approaches, including:

- Entity relationship (ER) models
- Resource Description Framework (RDF)
- Web Ontology Language (OWL)
- Unified Modeling Language (UML)

## What Kinds of Offerings in the Metadata Management Tool Market Can Help?

A broad range of offerings exists in the metadata management tool market. This market includes specific enterprise metadata management solutions, products with metadata management features and tools focused around big data. The tables below provide representative samples of metadata management tools, and are not intended to be all-inclusive lists.

Table 3 lists representative enterprise metadata management solutions.

# Internal Use Only

Table 3. Selected Enterprise Metadata Management Tool Vendors and Offerings

Vendor	Offering
<b>Adaptive</b>	<ul style="list-style-type: none"> <li>Adaptive Metadata Manager</li> </ul>
<b>Alation</b>	<ul style="list-style-type: none"> <li>Alation Data Catalog</li> </ul>
<b>ALEX Solutions</b>	<ul style="list-style-type: none"> <li>ALEX</li> </ul>
<b>ASG Technologies</b>	<ul style="list-style-type: none"> <li>Enterprise Data Intelligence</li> </ul>
<b>Cambridge Semantics</b>	<ul style="list-style-type: none"> <li>Anzo Smart Data Lake</li> </ul>
<b>Collibra</b>	<ul style="list-style-type: none"> <li>Collibra Data Governance Center</li> <li>Collibra Catalog</li> <li>Collibra Connect</li> </ul>
<b>Data Advantage Group</b>	<ul style="list-style-type: none"> <li>MetaCenter</li> </ul>
<b>DATUM</b>	<ul style="list-style-type: none"> <li>Information Value Management</li> </ul>
<b>Global IDs</b>	<ul style="list-style-type: none"> <li>Enterprise Information Management Software</li> </ul>
<b>IBM</b>	<ul style="list-style-type: none"> <li>InfoSphere Information Governance Catalog</li> </ul>
<b>Infogix</b>	<ul style="list-style-type: none"> <li>Data3Sixty</li> </ul>
<b>Informatica</b>	<ul style="list-style-type: none"> <li>Axon Data Governance</li> <li>Enterprise Data Catalog (EDC)</li> </ul>
<b>Oracle</b>	<ul style="list-style-type: none"> <li>Oracle Enterprise Metadata Management (OEMM)</li> </ul>
<b>SAP</b>	<ul style="list-style-type: none"> <li>SAP Information Steward</li> <li>SAP PowerDesigner</li> </ul>
<b>Smartlogic</b>	<ul style="list-style-type: none"> <li>Semaphore Cloud</li> </ul>

Source: Gartner (April 2018)

Table 4 lists representative products that include metadata management features.

# Internal Use Only

Table 4. Selected Products With Metadata Management Features

Vendor	Offering
Amazon Web Services (AWS)	<ul style="list-style-type: none"> <li>AWS Glue</li> </ul>
Cask	<ul style="list-style-type: none"> <li>Cask Data App Platform (CDAP)</li> </ul>
Cloudera	<ul style="list-style-type: none"> <li>Cloudera Navigator</li> </ul>
Datawatch	<ul style="list-style-type: none"> <li>Monarch</li> </ul>
Hortonworks	<ul style="list-style-type: none"> <li>Hortonworks Data Platform (HDP; Apache Atlas)</li> </ul>
Microsoft Azure	<ul style="list-style-type: none"> <li>Data Catalog</li> </ul>
OvalEdge	<ul style="list-style-type: none"> <li>Smart Catalog</li> </ul>
Paxata	<ul style="list-style-type: none"> <li>Paxata Adaptive Information Platform</li> </ul>
Podium	<ul style="list-style-type: none"> <li>Podium Data Marketplace</li> </ul>
Solix Technologies	<ul style="list-style-type: none"> <li>Solix Common Data Platform</li> </ul>
Talend	<ul style="list-style-type: none"> <li>Metadata Manager</li> </ul>
Tamr	<ul style="list-style-type: none"> <li>Tamr</li> </ul>
Teradata	<ul style="list-style-type: none"> <li>Kylo</li> </ul>
Trifacta	<ul style="list-style-type: none"> <li>Wrangler</li> </ul>
Unifi	<ul style="list-style-type: none"> <li>Unifi Data Platform</li> </ul>
Waterline Data	<ul style="list-style-type: none"> <li>Smart Data Catalog</li> </ul>
Zaloni	<ul style="list-style-type: none"> <li>Zaloni Data Platform</li> </ul>

Source: Gartner (April 2018)

## Recommendations

When planning your efforts to create and deploy an enterprise metadata management initiative, remember the three main components of such an undertaking: capture, categorize and use.

- Capture — Identify all data sources and engines for data movement and transformation as potential sources of metadata.** Such metadata is often generated whenever data is stored, extracted, transformed or transported — and the more broadly you can capture it, the more likely you will be to successfully manage, evaluate and gain business value from it. A complete

# Internal Use Only

metadata view enables lineage and advanced data analysis. Many organizations will already have systems creating metadata, and this metadata may be mostly ignored, or used only for one-off troubleshooting events. Data warehouses are a prime example of such systems, as they may be generating lots of metadata about ingestion, ETL and consumption functions. Multiple databases or data warehouses may create overlapping, but misaligned, schemas of metadata. Make an effort to capture all of these sources, ideally within an enterprise metadata management system, as this will create the most valuable view of data in the organization.

- **Categorize — Sort metadata into categories based on roles.** Role-based views of metadata help to clarify specific data management use cases — and to map the relevant metadata directly to these use cases. Such categorization will enable organizations to target analysis efforts and derive value from specific metadata threads. Existing sets of metadata sources may present conflicting or difficult-to-resolve data models. Aligning these sources with role-based views identifies overlaps in metadata and creates logical models to aid in metadata usage.
- **Use — Implement or improve metadata management solutions by linking metadata roles to specific business and operational processes.** This is the key to deriving business value from metadata. Once the metadata is categorized by roles and linked to processes, multiple useful business benefits can be derived from it. For example, holistic metadata lineage can enable EIM, MDM and data governance functions. Data catalogs and enterprise information catalogs are built from metadata. These catalogs enable consumer-based roles to effectively find, access and reuse data. A complete, end-to-end view of data, powered by a metadata-driven enterprise metadata management tool, also promotes newly discovered uses for, and correlations in, data.

## Conclusion

Metadata is being generated continuously — and with increasing velocity — by multiple sources and operations in the data pipeline. Organizations struggle to capture, categorize and use this metadata to support effective business and operational decisions. Existing metadata producers present one-off, isolated sets of metadata. Rarely do these metadata schemas align, making it difficult to standardize and resolve disparate sets of data. Enterprise metadata management solutions offer a complete, end-to-end view of metadata, supporting a harmonized presentation of these sources.

Data and analytics technical professionals responsible for enterprise information architectures and systems should employ a role-based approach to map metadata to specific use cases. This approach enables effective use of metadata to support EIM, data governance and critical business use cases. Data and enterprise information catalogs, built on enterprise metadata management systems, can reveal additional business insights.

Metadata management systems have traditionally focused on a *passive* view for metadata. Discovery, mapping and lineage are all well-developed functions in these systems. A more recent trend is toward *active* metadata management. Systems are now attempting to not only discover and categorize metadata, but also perform concrete actions. Such actions may include implementing

# Internal Use Only

data quality code or even executing business decisions based on changes in metadata. The value of enterprise metadata management solutions will increase with the maturity of these capabilities.

## Gartner Recommended Reading

*Some documents may not be available as part of your current Gartner subscription.*

"EIM 1.0: Setting Up Enterprise Information Management and Governance"

"Agile Data Quality to Maximize Your Business Results"

"Enabling Essential Data Governance for Successful Big Data Architecture Deployment"

"2018 Planning Guide for Data and Analytics"

"Magic Quadrant for Metadata Management Solutions"



# Internal Use Only

**GARTNER HEADQUARTERS****Corporate Headquarters**

56 Top Gallant Road  
Stamford, CT 06902-7700  
USA  
+1 203 964 0096

**Regional Headquarters**

AUSTRALIA  
BRAZIL  
JAPAN  
UNITED KINGDOM

For a complete list of worldwide locations,  
visit <http://www.gartner.com/technology/about.jsp>

---

© 2018 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. or its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. If you are authorized to access this publication, your use of it is subject to the [Gartner Usage Policy](#) posted on gartner.com. The information contained in this publication has been obtained from sources believed to be reliable. Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information and shall have no liability for errors, omissions or inadequacies in such information. This publication consists of the opinions of Gartner's research organization and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice. Although Gartner research may include a discussion of related legal issues, Gartner does not provide legal advice or services and its research should not be construed or used as such. Gartner is a public company, and its shareholders may include firms and funds that have financial interests in entities covered in Gartner research. Gartner's Board of Directors may include senior managers of these firms or funds. Gartner research is produced independently by its research organization without input or influence from these firms, funds or their managers. For further information on the independence and integrity of Gartner research, see "[Guiding Principles on Independence and Objectivity](#)."