Leverage the Metadata Management Maturity Model to Support Augmented Data Management

Published 1 November 2021 - ID G00751681 - 20 min read

By Analyst(s): Mark Beyer, Guido De Simoni, Melody Chien, Alan Dayley, Ankush Jain, Nina Showell

Initiatives: Data Management Solutions

Data and analytics leaders can use this maturity model to guide their organizations toward augmented data management utilizing advanced metadata techniques that increase efficiency while lowering design, development, deployment and maintenance costs.

Overview

Key Findings

- Metadata practices that focus exclusively on identifying, collecting and enabling search over static or passive metadata fall short in delivering business value.

- The majority of organizations struggle to break the pattern of metadata linked specifically to information delivery projects or initiatives, limiting reuse for internal data users and partners.

- Existing gaps in metadata force data science, artificial intelligence and advanced analytics users and developers to try to infer missing metadata for each individual effort.

Recommendations

Data and analytics leaders engaged in data management solutions initiatives must address active metadata and:

- Determine the gap between current state metadata management and the target state by cataloging currently available metadata and comparing the metadata inventory and its use to this maturity model.
Deliver advanced metadata management capabilities by focusing on priority data domains that cross business areas to assist in a phased approach to metadata-enabled augmented data management.

Evaluate team member skills and any tools in use to resolve the metadata management maturity gap between maturity Level 2 and Level 3 where appropriate.

Develop separate tactics for different metadata requirements by comparing active metadata requirements from broad enterprise to localized needs.

**Strategic Planning Assumptions**

Through 2024, passive data catalogs will be adopted by 60% or more of organizations but will provide significantly lower business value than expected.

Through 2024, active metadata capabilities will be embedded to facilitate cross-system orchestration across silos for new and existing systems, lowering human efforts by 40%.

By 2025, 20% of automated high-impact business decisions in an organization will leverage active metadata to enhance developing, monitoring and enhancing AI and decision engines while increasing trust.

**Introduction**

The purpose of metadata is to enable broader use and understanding of any data asset — and that applies to human utilization as well as augmented data management.

Metadata has remained partially useful and significantly unclear for many (if not most) organizations for the past 30 years. Metadata management has floundered in the churning environment of multiple system implementation projects, uncoordinated reporting and analytics use cases due to tight coupling with the systems being deployed. When you add to that the dissonant data quality and master data management efforts, it becomes even more difficult to manage.

This metadata maturity model provides guidance on how to execute on the journey from existing passive metadata practices toward an active metadata environment. This environment will enable AI-enhanced data discovery by utilizing appropriate machine learning, graph analysis and data inference to accelerate data utilization in the digital business.
Analysis

Data management solutions historically have relied on reactively collecting metadata as the best that most enterprises can do in most cases — almost to a point of being unaware that better solutions and practices exist. Improvements are found in converting from the passive utilization of any metadata that is available from using it “as is,” to an active metadata approach focused on both tactical and longer-term strategic organization business goals. This will significantly increase the digital capabilities of any organization. At the same time, this will lower data delivery costs while increasing data utilization in the enterprise.

Familiarity with any data is generally derived from using metadata that is often locked up in the minds of users, designers or developers. It is often the case that each individual system implementation has its own version of metadata with no effort taken to align those descriptors across multiple projects.

Then the debate begins about how much metadata and what types of metadata are even needed. This is also subjected to a filtering effect that favors immediate, local requirements.

The term “metadata” describes various facets of a data asset used to improve its usability throughout its life cycle (see Note 1). Metadata management is about an organization’s management of its data and information assets in order to address use cases such as data governance, risk and compliance, analytics, and data monetization. Business benefits that motivate the introduction of a metadata management solution to an enterprise include:

- Reuse of data for ease of sharing
- Improved productivity of business functions from finance to operations
- Clear identification of data assets used in regulatory roles to enable the avoidance of fines, fees, penalties and even imprisonment
- Improved risk management and better assessment by decision makers of the impact of change within an enterprise
- Reduction of technical debt
- Systemic coherence across applications and models consuming the metadata
- Assessment of data content relative to business purpose for quality metrics
Determine the Gap Between Current State Metadata Management and the Target State

Most organizations have some form of limited enterprise metadata programs, but most organizations rarely achieve any form of closed loop between various data management projects required to enable enterprise metadata management. Isolated efforts to link individual projects and their metadata generally result in creating domains in which competing metadata practices and standards are incompatible when considered together. However, metadata is effectively a graph, no matter how it is structured, and can be resolved if taken out of specific deployment context. Yet if any grand, organizationwide approach is attempted from the center out, it usually results in rejection at the project level — and even rebellion that intentionally creates an isolation effect to protect timelines and budgets.

Figure 1: Metadata Management Maturity Model
The expectations at each level of maturity gradually increase in both amount of metadata and its utilization. Importantly, it is entirely likely that different business units, delivery channels, and even department- and team-level operations will have different maturity levels, even in adjacent operations. Further, within the same logical cluster of activity, some processes within a single system may have a much higher maturity level than different parts of the exact same system. Humans with varying skill levels will require different types of tools and even interfaces for different human users with the same tools or platforms. Different types of user interfaces and workflows will be required to ensure user adoption and utilization of any metadata management and metadata analysis tools.

**Metadata Management Adoption Is Accelerating for All Approaches**

A broad variety of new use cases is driving investment in metadata management. It is also suggested by interactions with vendors. These investments point to new use cases for older platforms like data warehouse modernization. They also highlight newer demands like streaming analytics and those that mix long-held practices with new technology approaches such as “360 degree” solutions (for customer, reference, product and supplier data), MDM, explainable AI and DataOps.

As a metadata management approach matures, it will become increasingly important to adapt how it is organized and where responsibility is assigned for it in the broader enterprise. The drivers for this metadata adoption are:

- The rise of semantic formalism (also known as formal ontologies) for improved interoperability
- The transfer of metadata ownership from the CIO to the chief data officer (CDO) or a similar role
- The increase in the variety and extent of metadata supported
- The enhancement of the scope of metadata through automation (ML) and through automated enrichment by semantic search capabilities, standard processes and crowdsourcing
- The development of shared understanding across multiple domains
- New ways to capture and visualize metadata (driven by data preparation for analytics)
Most organizations will find that their current metadata management practices differ across applications, data and technologies, and that these practices are siloed by the needs of different disciplines — each with its own governance authority, practices and capabilities.

Over the period starting in mid-2021 through late 2028 and beyond, metadata management will finally begin demonstrating consistent value through emergent and improved tools and platforms. Organizations will also realize new strategic value that will become invaluable for combining data integration, quality, governance, mastering and even database designs.

Active metadata management will drive these currently disparate practices and technology solutions toward a comprehensive, cohesive and fully complementary data management backplane. This new approach will resolve many of the issues relative to distributed data management and utilization, leverage the full capabilities of the cloud, and displace and disrupt practices relative to any form of “hybrid” data management forever. The separation of passive practices from active, and of active practices into delivery channels of “branded” or “white label” (see Note 2) will all be supported by the most mature metadata management tools.
Figure 2: Relative Adoption of Metadata Approach

Relative Adoption of Metadata Approach

- Passive Metadata
- Active Branded
- Active White-Label

Source: Gartner
Through the next decade, metadata management that focuses on passive techniques such as catalogs that collect searchable metadata, data management tools that utilize design metadata (e.g., databases, integration, quality and application platforms) and runtime metadata from systems will continue to increase in adoption. At the same time, adoption of active metadata will increase, either through branded platforms or supporting technology that is embedded in adjacent data management tools (“white label”). Branded technology will lead the active metadata adoption trend at first, but then be overtaken by embedded metadata interfaces and metadata sharing and analysis.
In parallel, organizations pursuing passive metadata practices will see they cannot achieve 100% effectiveness relative to the current metadata capabilities. However, when they add active metadata strategies, they will exceed the current anticipated value — setting a new expectation level as active metadata approaches mature and are deployed — ultimately achieving more than double the value currently anticipated from popular passive metadata strategies.

**Evaluate Team Member Skills and Any Tools in Use to Resolve the Metadata Management Maturity Gap**

Some tech vendors enable a technology-driven delivery approach to metadata management, which leads to a wide usability and suitability gap for the ultimate users. However, the technology is not often used to its full potential.

The potential for augmentation and machine learning (ML) is not yet realized, even while these expectations in the market are accelerating. AI augmentation of all aspects of data and analytics is making everyone in the content creation pipeline more productive with less skill and changing how and where more types of users interact with analytic content.

Finding relationships in combinations of diverse data by using graph techniques at scale will form the foundation of modern data and analytics. Metadata is the key for almost all forms of automation and augmentation in modern data utilization scenarios.

Metadata has historically been considered highly technical in nature, describing schema, schema rules and how data is represented in any given dataset relative to data expectations. However, even technical metadata such as capacity and utilization of the processing workloads can enhance system optimization using advanced ML and augmented data management, or it can be used for augmented analytics. Metadata can also be used to prove that compliance requirements have been met.

Metadata analysis will determine the best locations, access pathways, rights management and processing demand. It will even control how multicloud environments coordinate capacity and utilization, resulting in better processing parameters and resource allocation. Highly emergent workloads will be dynamically deployed across increasingly broader infrastructure that will eventually disregard all specific provisioning of assets in favor of policy (including budget policy for “augmented FinOps”) as well as demand-driven deployment models.
Some modern metadata management solutions can track the activities of data consumers in order to understand actual data usage, what data is most important, which datasets are related and the nature of those relationships. These solutions also incorporate data and query usage analytics and certification metrics. Their capabilities help business users identify which integrated datasets have been certified, or not, for use by information stewards, as well as for use by experts in business processes and analytics. They can also use this information to optimize production delivery of that data using frequency of access as a proxy for value to identify the best ROI.

Develop Separate Tactics for Different Metadata Requirements

Enterprise metadata initiatives are often distracted by a false dichotomy between business and technical metadata. In fact, metadata is just metadata, and it can inform the technical user and the business users equally. For example, showing how many records flow through a data quality process and how many corrections are made and what data is rejected to a “bit bucket” is useful for both business and technology professionals.

Technical metadata that describes data structures, places some limits on content, provides some clarity around validation rules and indicates locations and access rights frequently requires manual analysis to leverage it. As a result, usage is usually limited to database designers, application designers, integration experts and limited end users.

So-called “business metadata” often consists of competing glossary terms and basic descriptions of business processes. Business metadata frequently mistakenly identifies use cases that do the “first capture” (mistakenly referred to as data “creation”). This is intended to separate first use cases from those that reuse data. The resulting assumption is that this process connects business tasks.

Importantly, the source of this consternation is generally found in how the same data is often developed and then reused in new structures and formats through a project-based implementation and delivery model. All of these activities begin with designs and specifications, then eventually start producing logs. All of that is metadata that can be used to help identify critical domains for data that is frequently reused, but often in variable and isolated forms and structure.
The Metadata Management Maturity Components

Utilize Table 1 to assess current maturity levels, but also to develop a roadmap for increasing metadata benefits in the organization. It will be important to prioritize different areas to focus on as your organization progresses through the maturity model. Also, some parts of the organization will be served perfectly well by lower maturity levels — it is not an absolute requirement that every part of the enterprise achieve Level 5.
Table 1: Metadata Management Maturity Levels
(Enlarged table in Appendix)

<table>
<thead>
<tr>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gartner client interaction and inquiry data from 2020 through 2021</td>
</tr>
<tr>
<td>Data practitioner interviews November 2019 to July 2021</td>
</tr>
<tr>
<td>CDO Agenda 2021: Influence and Impact of Successful CDOs in the Sixth Annual CDO Survey</td>
</tr>
<tr>
<td>Gartner secondary research services reviews of published case studies, press releases and available website references as of July 2021</td>
</tr>
</tbody>
</table>
Gartner for Technology Service Professionals research for market share and market forecasts as of July 2021

Multiple vendor/provider service briefings

Gartner Peer Insights client product reviews and ratings

**Note 1: Metadata Maturity Can Be Inconsistent**

The market for metadata management solutions continues to evolve (see The State of Metadata Management: Data Management Solutions Must Become Augmented Metadata Platforms).

The market demand for ever-increasing capabilities from metadata management solutions sees organizations pressing for more innovation and significantly improved support and skills. Metadata management solutions have improved over time, but still have gaps in workflow, data lineage, source data connectors and augmented catalog capabilities (see Market Guide for Active Metadata Management).

Many organizations are just starting to explore enterprise metadata management solutions as a mechanism to optimize their data utilization and management and to improve productivity.

At the same time, natural language technology solutions have done much to accelerate the harvesting and representation of metadata from unfamiliar structures like content, email, documents and even audio-to-text analytics.

**Note 2: What Is “White Labeled” Active Metadata Delivery?**

When active metadata becomes the primary focus of metadata management and utilization, there will be two channels for delivery that will ultimately be supported by metadata solutions.

- Branded metadata management will consist of named platforms or tools as products that collect, analyze, develop recommendations from and even support orchestration with metadata. These solutions will collect metadata from third-party tools and ingest it into internal graph databases and graph analytics solutions to accomplish the goal of active metadata.
"White label" metadata management will perform all of the same functions as a branded metadata management solution. However, they will specifically support the concept of deploying distributed agents, services or daemons that run “inside” third-party tools and communicate their analysis findings, develop ensemble-based recommendations and execute orchestration almost as “plug-ins” to those tools. These are called white label because the branding “disappears” inside the other data management tools.

Importantly, the same provider can offer both delivery channels. Some will offer them as either/or options that are selected by type of deployment desired. Ultimately, the successful vendors will be able to support both delivery channels interchangeably and even include internal analysis to determine which delivery channel is appropriate based on the requirements of specific data management environments.

**Note 3: Metadata Works With All of Data Management**

There will be a “multiagent data fabric” in the future. Data entities (and data producers and consumers) being represented by autonomous agents that proactively build their own metadata will “roam” the fabric to see if they can “attach” or align themselves with other agents by (fuzzy) matching each metadata. By learning, each agent continues to evolve its own metadata.

In this case, the form of the data fabric would be a fabric where data entities/agents would be rewarded by finding a match with others, either by being consumed or through other feedback. Other artificial/human actors could also be active in such a fabric. These could include privacy protection agents, security agents, broker agents and so on. This would also allow the application of reinforcement learning to enforce successful metadata strategies at entity level, or other higher level (e.g., integration strategies).

---

**Recommended by the Authors**

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

- **Market Guide for Active Metadata Management**
- **The State of Metadata Management: Data Management Solutions Must Become Augmented Metadata Platforms**
- **Quick Answer: Key Metadata Capabilities for Data-Driven Organizations**
Emerging Market Scenarios Are Driving a Renewed Focus on the Capabilities of Metadata Management Solutions

Accelerate Adoption of Enterprise Metadata Management Solutions by Improving Usability for the Business

Solution Comparison for 7 Data Fabric Offerings
### Table 1: Metadata Management Maturity Levels

<table>
<thead>
<tr>
<th>Scope</th>
<th>0. Project</th>
<th>1. Inventory</th>
<th>2. Catalog</th>
<th>3. Preactive</th>
<th>4. Active</th>
<th>5. Augmented</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individual metadata assets are defined for applications, databases, integration and mastering and largely remain local to projects or deployments relative to their extent and persistence.</td>
<td>Colocation of existing metadata and materials. Adds to business user knowledge regarding current assets available.</td>
<td>Retention of data inventory, aligned assets, user interface for search/identification, beginning of metadata maintenance roles. Utilization of import/export metadata and system integration approaches to combine diverse metadata. Begins tracking social network of data users to accelerate collaboration in data usage.</td>
<td>Introduction of analytics capability to evaluate and resolve metadata conflicts. Resolution of technical data taxonomies required. Identification of patterns or trends, origination, and responsibility for metadata changes and updates.</td>
<td>Convert existing metadata knowledge and tools from human-driven activity by separating repetitive or mathematics-capable analysis into ongoing data profiling, machine learning over log analysis then combining both with existing passive metadata for graph analytics. Targeted reduction of manual delivery while increasing &quot;push&quot; model to business users notifying them of</td>
<td>Combination of human and automated metadata analytics to enable automated processes to run independently of human instruction or action. Can also identify situations in which the decision base is inadequate and requires human intervention (identify unacceptable levels of &quot;force fitting&quot; of the analysis). Metadata begins to push its findings to users and communities instead of relying on human-driven Search.</td>
</tr>
</tbody>
</table>

Gartner, Inc. | G00751681 | Page 1A of 6A
<table>
<thead>
<tr>
<th>Metrics</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accumulated counts and descriptions are present. Generated list of metadata gaps up to and including &quot;no metadata available&quot; logging.</td>
<td>Report the ratio of assets in catalog compared to assets assumed in organization. Identify the percent of assets with assigned stewards. Improve the frequency of metadata updates. Report and analyze the lag time between metadata source changes and catalog entity updates. Report on the Frequency and distribution of</td>
</tr>
<tr>
<td>Count of individually developed data schemas resolved into a given taxonomy. Taxonomic domains/subjects resolved. Rate of absorption of systems into existing taxonomies. Rate of taxonomic domain/subjects deployed notably accelerates.</td>
<td>Ability to count reused data objects with focus on thresholds for &quot;most common&quot; use (inertia). Capability to discern frequently linked/used data objects by density and number of connections to reports, analysis, business processes and applications (gravity). Monitoring processing requirements for differing</td>
</tr>
<tr>
<td>Identification of and count of AI decision engines built that utilize thresholds determined for inertia, gravity, mass and sympathy of data assets as decision inputs for recommendations. Comparative analysis of level of human effort from previous projects from project delivery to current maintenance levels against any recently deployed augmented data management</td>
<td></td>
</tr>
</tbody>
</table>
different roles of users. combinations of data (mass). Trend analytics of each of the above and graph analysis of how their components interact (sympathy).

systems with similar mass, inertia, gravity, and sympathy (MIGS) across multiple use cases. Utilization trend analysis of any deployed orchestration utilizing automated data management and rate of acceleration/deceleration of each.

Governance policy reporting, enforcement and evolution

Minimum schema and business purpose descriptions

All metadata assets taken “as is.” Missing metadata developed as needed without specific retention when completed.

Stewards assigned to metadata domains by asset type or source origin. Development of standards for metadata submissions to the governance board relative to structure and completeness.

Introducing multiple definitions in the glossary with business unit, process channel or vertical/horizontal practitioner terms as needed. Utilizing underlying data taxonomy, the glossary begins to identify data affinity

Identify rates of changes in metadata patterns and introduce alerting and workflows. For implementation, the workflows can be human, augmented or automated based on the persistence of

Controls on data lakes, data warehouses, data marts, data hubs, data services and any other form of data reuse/sharing backplane to leverage existing processes when the same data assets
to help with synonym resolution and exclusion of homonyms or near-homonyms (e.g., are prospect, client and customer the same or not?)

the metadata models and behavioral changes.

exhibit similar MIGS characteristics. Decision model as to when new services/processes are required based on metrics thresholds to control devolution patterns. Governance and stewardship are as automated as possible.

<p>| Organization | The project manager/designer determines the entire extent of metadata required based on completing the project or delivery. | The demand for a data inventory, usually in response to application or data platform design; or, minimal regulatory requirements. | Data stewardship role established by position or assigned to existing roles. Reporting to enterprise architecture team or CDO (if there is a nominal CDO function regardless of title). | Beginning of data engineering supported at any emergent area (e.g., data science, data integration, master data analysis). Transfer of metadata ownership from the CIO to the chief data officer (CDO) or a similar role. Demands new roles in business | Coordinated leadership across data repositories/databases, data integration, data quality and data mastering for the purpose of either registering or accumulating metadata. Inclusion of application | Independent development teams are no longer required to develop data capture based on canonical or common model specifics due to recognition of data content analysis capabilities of automated functions. |</p>
<table>
<thead>
<tr>
<th>Process</th>
<th>Metadata is embedded within design documents, custom-code or development tools.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Introduction of formalized glossary, definitions and schema tracking. Scheduled updates to catalog entries based on metadata import/export functions. Initiation of “data integration” projects and practices for integrating metadata in a similar fashion as well as enabling analysis of metadata patterns and trends.</td>
</tr>
<tr>
<td></td>
<td>Introduction of scheduled metadata updates, which can be a periodic human effort or event/schedule driven updates from/through data management tools. Practices for using data use logs in analysis/design established.</td>
</tr>
<tr>
<td></td>
<td>Data management log analytics to determine usage and utilization characteristics of enterprise data assets to build an analysis graph. Introduction of metadata regarding data profiles, content analytics and schema introspection. Orchestration instructions or job cards issued from the metadata recommendations to.</td>
</tr>
<tr>
<td>Technology</td>
<td>Diverse design interfaces and physical diagrams.</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Spreadsheets. Basic content templates. Rudimentary data catalog functions.</td>
</tr>
<tr>
<td></td>
<td>Data catalog platform or implementation of custom database for metadata capture/store and an interface that facilitates the scope, governance and process functions as indicated.</td>
</tr>
<tr>
<td></td>
<td>Import/export of metadata from design tools for integration, quality, mastering, stewardship, to/from metadata location. Identification of in-stack or independent metadata analysis capabilities for trending, coincidental and common use cases for different data objects.</td>
</tr>
<tr>
<td></td>
<td>Machine learning to perform gravity, inertia, mass and sympathy analysis. Machine learning to identify anomaly patterns within trends in metadata changes. Content analysis tools to infer missing metadata and evaluate content change patterns. Data profiling tools to infer missing metadata and data value change patterns.</td>
</tr>
</tbody>
</table>

Source: Gartner