Quick Answer: Key Metadata Capabilities for Data-Driven Organizations

Published 16 August 2021 - ID G00753531 - 5 min read

Nina Showell

Initiatives: Data Management Solutions; Data and Analytics Strategies

Metadata fuels data-driven organizations by providing visibility and context around data and its characteristics, such as location and use. To become more data-driven, data and analytics leaders must engage with the business to build the five metadata capabilities discussed here.

Quick Answer

What metadata capabilities are key to becoming data-driven, and how do I build them?

1. **Data semantics**: Define data at the right level for business users by providing clarity around data definitions and variations across systems, departments and use cases.

2. **Data location**: Implement end-to-end observability into where data sits. Reveal the provenance/system of record, the other data copies (including in downstream systems), the lineage and the data quality enrichments across the various systems.

3. **Data access**: Provide visibility into the chain of custody (data owners, data consumers and boundaries), and ensure auditability of who has access to what data.

4. **Data trust**: Provide transparency on data quality issues, compliance requirements, stewardship efforts and data literacy efforts.

5. **Data utilization**: Measure data usage (actual data experiences) and enable data sharing, because usage frequency indicates which data has high business value.
More Detail

In their quest to become data-driven, organizations push two essential changes:

1. Make data people business-aware
2. Make business people data-driven

To meet these goals, data and analytics leaders must engage with their business counterparts to build five metadata capabilities, as illustrated in Figure 1.

Figure 1: Metadata Capability Framework

Building Your Metadata Capabilities

1. Data Semantics — “How Do I Define Data at Usable Levels?”
“Semantics” refers to definitions/meanings. Data semantics should be defined and managed at the level/understanding of common business users who need data for their day-to-day use.

**Challenge:** Practitioners have been applying software engineering principles, such as abstraction and reusability, to data definitions and models. For example, they’ve been defining suppliers, customers, prospects and partners under a unified “party” domain. However, these terms mean different things to different consumers. Such normalization efforts skew data definitions from usable levels.

**Recommendations:** *Define data at the right level of use.* Enable all consumer personas to easily find data in their context, understand data definitions (along with any exceptions/variations), and contribute to the semantics, so that they can consume data with assurance.

2. **Data Location — “How Do I Improve Data Location Observability?”**

Observability involves knowing the provenance of data, the data capture channels, the other systems that hold a copy of the data, and the changes (such as quality and chain of custody) between the primary and secondary copies of data.

**Challenge:** Organizations lack end-to-end observability to comprehend the true data impact. Also, data lineage capabilities are limited to supporting technology.

**Recommendations:** *Observe the creation, transformation and application of data across locations throughout its full life cycle.* Let data consumers pick the preferred location for their use case — say, fixing incorrect customer data at the upstream source, or scoring customer churn propensity at a downstream system.

3. **Data Access — “How Do I Simplify Data Access?”**

The objective here is to democratize data and enable self-service among business users.

**Challenge:** The fragmented chain of custody across systems is the foremost challenge in democratizing data for consumers. Furthermore, attempts to collect all data are not practical when that data is constantly growing and changing.
Recommendations: Make the chain of custody visible by delineating data owners and data subject boundaries. Classify data by information protection levels, and label it by domains. Study data usage patterns to inform policies around which data can go where and who can/should access it. Minimize data duplication. Use a mix of data integration approaches — spanning from physical to virtual delivery, and from batch to event-driven data propagation.

“We make all data accessible for our colleagues by default, and apply restrictions only where necessary.”

— Olli Voima, Data Value Officer, Orion

4. Data Trust — “How Do I Improve Data Trustworthiness?”

Trust is a bigger issue than what metadata alone can solve. However, data trust is a critical metadata capability for becoming data-driven. It focuses on the line of balance between data readiness and decision making, addressing questions like:

- Why aren’t these numbers matching?
- What action does this data warrant?
- Why is this data different from my expectations?

Challenge: Data consumers may not trust data for several reasons, such as integrity issues, poor quality, lack of interest/awareness around data stewardship efforts, or even presumptions.

Recommendations: Drive data literacy efforts with social metadata like user feedback, developer feedback and frequency of access. From running regular show-and-tell sessions to driving enterprisewide data literacy programs, maximize business collaboration and education channels when addressing data issues related to semantics, location, access and trust.

5. Data Utilization — “How Do I Improve Data Use and Reuse?”

Data usage is a tangible metric of the impact that data creates in an organization. It gives an indication of the operational value realized from data.
**Challenge:** Organizations do not study their runtime metadata (database query logs, integration job logs, data quality audits, etc.) sufficiently enough to challenge the status quo around their design-based metadata (data models, schema definitions, glossaries, etc.).

**Recommendations:** Start your metadata discovery journey with runtime metadata. Apply active metadata practices by collecting and analyzing all forms of metadata. Feed the results to machine learning models that produce recommendations and automation metrics as output (see What Is Data Fabric Design?). For instance, to determine alignment between "data as designed" and "actual data experiences," you must continuously analyze data usage across systems and users. Share data usage metrics with business users to drive further usability.

**Addressing Operational Challenges**

Staying up to date with metadata is a major challenge. There are two operational challenges you must address when implementing the metadata capability framework described above:

1. Automate the operational tasks related to the metadata capabilities. Let the data fabric design guide your near-term technology investment decisions (augmented catalog, knowledge graph, etc.).

2. Compile the necessary skills. Successful implementations require several competencies, including business acumen, technical know-how, domain understanding and governance. Find the right balance of these competencies, and incentivize your top contributors to keep them motivated.

**Recommended by the Authors**

- Data and Analytics Essentials: Metadata Management
- 5 Ways to Use Metadata Management to Deliver Business Value From Data
- The State of Metadata Management: Data Management Solutions Must Become Augmented Metadata Platforms
- What Is Data Fabric Design?
- How to Use Semantics to Drive the Business Value of Your Data