DIGITAL DIALOGUE

THE MANY FACES OF METADATA MANAGEMENT—FROM DATA CATALOGS TO DATA GOVERNANCE AND BI OPERATIONS

Metadata is "data about the data," and managing it correctly can improve data transparency and accessibility while helping organizations track and understand data lineage and improve data governance. Good metadata also gives users more confidence in BI reports and analytics.

Unfortunately, TDWI research finds that most organizations still acquire and manage metadata manually and haphazardly using methods that are slow and do not scale. In addition, there is confusion about the differences between metadata repositories, data catalogs, and business glossaries.

The good news is that understanding and applying better metadata management to your organization's data management and integration practices can help with BI reporting, analytics, data governance, migration, and regulatory adherence.

tdwi
TRANSFORMING DATA WITH INTELLIGENCE™

WHERE METADATA MANAGEMENT IS HEADED

We are in a period of rapid change in how data is managed and shared within companies, including trends such as the democratization of data and self-service BI. As more users throughout an organization seek to interact with data and to personalize their interactions, "it opens up the potential for data chaos."

That's according to David Stodder, a TDWI senior research director who focuses on business intelligence. In a recent webinar, Stodder described how organizations often struggle with new models for quick, agile data manipulation and access.

To facilitate faster development and deployment of tools that give users better access to data, organizations need easier ways to integrate new data and to change views, Stodder explained. From carefully defined and structured data models, "we're moving into a world where there's going to be a lot of growth and new kinds of data, [all] coming in very quickly through this digital transformation that so many organizations are going through."

Organizations are also seeing a move toward multiplatform architectures, he noted, including cloud use and data lakes—and are looking for ways to knit together these often-disparate systems.

In the face of all this, metadata management offers a way to better control, identify, and track disparate pieces and types of data across platforms and locations.

"Metadata is information about the data, and it allows organizations to begin to describe data and match it up with other kinds of data," Stodder said. Creating a set of identifying data helps with searching for, finding, and organizing the data, as well as relating to it, which is critical for analytics.

METADATA MANAGEMENT, DATA CATALOGS, AND GLOSSARIES

Key to understanding metadata is to understand the differences between metadata repositories, business glossaries, and data catalogs. TDWI research indicates confusion about these terms, how they differ, and how are they applied to different use cases. "Metadata management and data catalogs are both hot concepts right now," Stodder said in the webinar, "and there's some confusion in the industry about how they fit together."

Metadata management, he explained, focuses on understanding the data and what it is, then using that understanding to properly update and expand reports, dashboards, and other data artifacts. A data catalog, in contrast, is a store of information about an organization's data assets. It gives users a mechanism for identifying disparate data sources that can be used together to extract useful intelligence to act on. Stodder has written that "an up-to-date, comprehensive data catalog can make it easier for users to collaborate on data because it offers agreed-upon data definitions they can use to organize related data and build analytics models."

Most data catalogs get metadata from databases, but a complete data catalog also requires metadata from reporting tools—which is called a business glossary. A data catalog requires a business glossary; however, with many data catalogs in place, it is challenging to extract metadata from across the organization. Therefore, tools that can automate creation of the business glossary, as well as manage data catalogs and metadata, are becoming increasingly popular.

Stodder sees data catalogs and metadata as complimentary: "If you have metadata management, it improves your data catalog because you have metadata as the core of the data catalog." In order for your data

catalog to be effective and efficient, it's important to be able to manage that metadata so it can do a good job of indexing the data.

One common issue with data catalogs is that different systems within an organization have their own data catalogs. This is often true after mergers or acquisitions, for example, and points to the need for some kind of universal system to tie together data catalogs and business glossaries—a central repository where metadata is stored and managed. There, a metadata management tool can act as a kind of "über-data manager," Stodder said—a universal tie that can be used to govern and manage data across all data catalogs, repositories, and business glossaries.

BUSINESS DRIVERS BEHIND METADATA MANAGEMENT

Good metadata management can improve BI reporting, analytics, and regulatory reporting. One example Stodder offered is audits. When an organization is audited by an outside agency, business and technology leaders within the company must respond quickly.

Often, that requires teams to manually pull together the necessary metadata. It can be an incredibly difficult process, Stodder pointed out, because IT and subject-matter experts take time away from revenue-producing activities to find the needed data and document it to meet audit requirements.

The more this process can be standardized ahead of time, the better—and metadata management and data preparation tie into that idea. Data preparation is a necessary and ongoing process in servicing dashboards, analytics, and other BI activities, Stodder said, but metadata, too, must be "right there in the middle of data prep, really at almost every step."

That makes it critical to avoid delays and bottlenecks in metadata management because they slow data preparation as well. In fact, TDWI research shows that when it comes to recent BI and analytics projects, respondents reported spending up to 80 percent of their time preparing the data.

"The more organizations can bring this time down, the better," Stodder said, as this allows users, data scientists, and analysts to spend more of their time analyzing the data and applying those insights to business problems.

Again, centralized metadata is important here, he emphasized: "It's important to begin to understand the data at the sources—and then to understand and not just improve processes but to be able to redo them." That means being able to look back, particularly for audit and data lineage purposes, and understand what happened to the data.

All of that, Stodder said, can be recorded in a good metadata management system, which makes metadata available from all stages of the pipeline.

THE NEED FOR GOVERNANCE AND TRUSTED DATA

Increasing numbers of regulations, such as the European Union's General Data Protection Regulation (GDPR), require companies to be able to prove they can control the entire data movement process.

That can be challenging, as described here, with various siloed systems across the enterprise, combinations of cloud and on-premises data, and rapidly changing environments and user needs. Data managers need to visualize and understand the complete data flow in a changing environment. For these (and other) reasons, good metadata management is critical to enforcing data governance.

Metadata management is clearly key to knowing what data the organization has. It's also critical to establishing data lineage and knowing what was done to the data over time, as well as knowing the roles and relationships of users working with the data. In that sense, metadata management is essential to improving trust in the data.

It also plays a central role in data governance and regulatory adherence. Metadata in the data catalog determines what a user is allowed to see, for example. Users may be viewing analytics results in a dashboard or another kind of report, or looking at data models. It's important for that user—whether a data scientist, analyst, or other business or technical user—to know what was done with the data before it reached this point.

That transparency goes to the issue of trusted data. "It's critical that we can understand whether we can trust what we're seeing" as users, Stodder said. Again, metadata serves as a way to establish what was done with the data and to trace and preserve the roles of users who touched it and what they did.

All of this builds up trust in the data and underlines the importance of the central role of metadata in data governance and regulatory adherence across the organization.

RECOMMENDATIONS

For better metadata management, Stodder recommended addressing three areas.

First, technology leaders should evaluate new technologies for metadata management that can make it easier to find and manage data and BI reports. They

Octopai and the Many Faces of Metadata Management

As a metadata management automation company, Octopai offers discovery and lineage tools not only for managing data catalogs across the organization but for BI operations and data governance as well. The software acts as a single cross-platform product that centralizes the organization's metadata in one place, enabling visibility and control of metadata that is typically scattered across the enterprise—all without tedious and time-consuming manual efforts.

- Using a sophisticated search engine with hundreds of crawlers, Octopai searches the BI landscape and delivers a 360-degree view of metadata within seconds.
- Metadata is automatically gathered from a wide variety of sources, including ETL, databases, and reporting tools. It is stored and managed in a central repository.
- Smart algorithms can model and index all metadata types for quick location and understanding of cross connections.
- Octopai creates a full lineage of the data journey as it flows through multivendor systems and can search for reports or references and present the complete data flow in seconds.
- As an out-of-the-box, cloud-based product, it can be up and running within a day. There's no installation, no special training, and no organizational process changes required.
- Running a proof-of-concept in Octopai is free, fast, and easy to create. Details are available at www.octopai.com.

can then use metadata knowledge to increase data lineage understanding and apply metadata management to data movement and migration.

Second, it's critical to bring business stakeholders into the definition process and to address business objectives as part of bringing business on board regarding the need for metadata management tools. Focus on how new technologies can be helpful in meeting business objectives such as understanding customers better, improving risk management, and addressing regulatory issues and audits in a more proactive manner.

"It's important to look at new technologies that can improve the ability to meet particular objectives," Stodder stressed. Also, it can help to point out that easy-to-use, automated tools will help users and developers work faster.

Finally, apply modern metadata management to improve governance and trust, Stodder said. This applies across platforms, including data lakes and the cloud. Data stewards can use smart automation to improve analysis and tracking of data selection, lineage, and auditing.

"A typical business user looking at a dashboard may not understand that some of the data is of a higher quality than other data," Stodder said. IT leaders can address that issue by using metadata knowledge not only to improve quality procedures but also to help users understand the level of quality, completeness, and consistency in the data.

In summary, Stodder reminded attendees that "metadata has a role in all of the different steps that are important to handling data governance in organizations and contributing to it." He encouraged the use of smart automation and good metadata management tools to improve the analysis and tracking of data across the enterprise, from data catalogs to data governance to BI operations and more.