International Journal of Engineering, Science and Mathematics
Vol.8Issue 9, September2019,
ISSN: 2320-0294 Impact Factor: 7.875
Journal Homepage: http://www.ijmra.us, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Implementing a Successful Enterprise Metadata Management Program

Sowmya Tejha Kandregula | Data Governance Leader
Brambleton, Virginia | sowmyateja@gmail.com

Abstract

An effective metadata management program is a good starting point to catalog an enterprise's data assets through a scalable and automated framework to support data catalogs, glossary (and) dictionary solutions.

This helps in promoting effective data management, data discovery, democratization, and standardization.

Keywords

Data Management, Data Governance Office, Critical Data, Data Custodian, Enterprise Metadata Collection, Metadata Management

Introduction

Metadata describes the data itself (e.g., databases, data elements, data models), the concepts the data represents (e.g., business processes, business terms, reports), and the connections (relationships) between the data and concepts. Maintaining a single source of certainty for data assets—like what it is named, what it means and how it is related to other assets—is critical to treating data as a strategic asset.

Metadata offers the basis to describe the content, context, quality and structure of data. It contains information needed to understand and effectively use the data. Metadata management helps successfully increase business value and reduce risk by capturing and governing information that empower stakeholders to find, understand and share trusted data.

Typically, a Data Governance Enablement team helps business users to confidently answer three fundamental questions:

- What data is available?
- Where did it come from?
- What is the level of trust?

Definition of Metadata

Metadata is all the information about an asset that allows data stewards to find, understand and make decisions on data by treating it as an enterprise strategic asset.

Metadata describes the data itself (e.g., databases, data elements, data models), the concepts the data represents (e.g., business processes, business terms, reports), and the connections (relationships) between the data and concepts.

Metadata Types

- Business Metadata
- Technical Metadata
- Operational Metadata

Metadata Format

Metadata contains the following information about a data element:

- Name and description
- Related business term and data domain
- Ownership and governance policies applied to the asset
- Data Classification
- Valid values
- Business rules
- Data Quality rules and metrics
- Audit trail
- Performance/Operational Metrics
- Lineage

Metadata Framework

Metadata Drivers

Metadata is critical to operate an efficient Data Governance and Management program. Below are the expected outcomes of a successful metadata program.

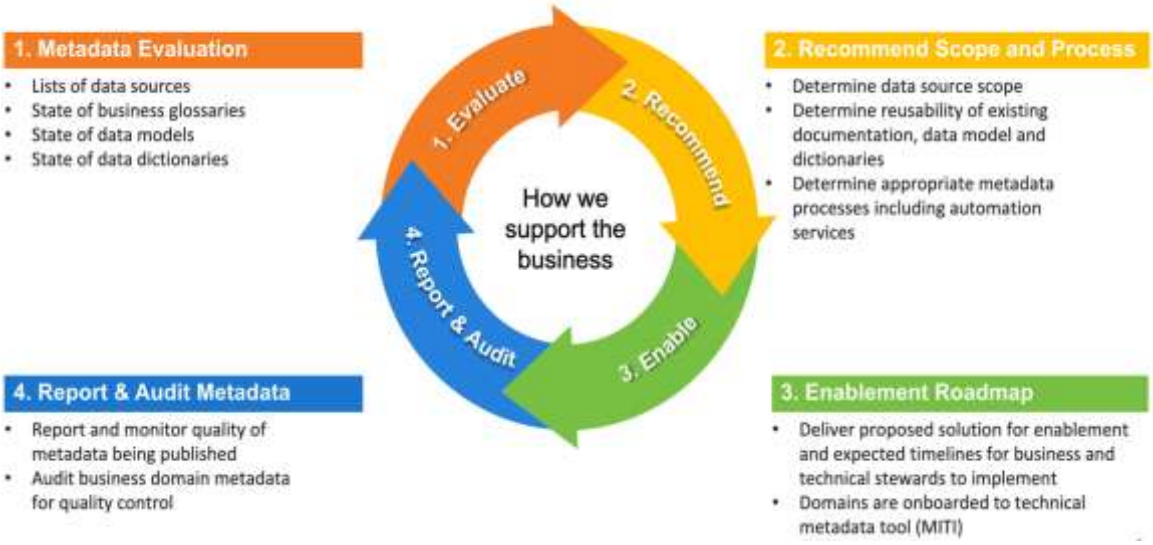
- Increase confidence in data by providing context and enabling the measurement of data quality
- Increase the value of strategic information by enabling multiple uses
- Improve operational efficiency by identifying redundant data and processes
- Prevent the use of out-of-date or incorrect data
- Reduce data-oriented research time
- Improve communication between data consumers and IT professionals
- Create accurate impact analysis thus reducing the risk of project failure Improve time-to-market by reducing system development life-cycle time
- Reduce training costs and lower the impact of staff turnover through documentation of data context, history, and origin
- Support legal, privacy, governance, and audit compliance

Metadata Enablers

Data Dictionary	Instantly locate metadata from multiple systems, so you can better organize and manage data.
Data Lineage	Get a complete view of your entire data journey, with visibility into full data context, from source to target.
Business Glossary	Create and manage glossaries to ensure consistency in data assets across organizations.

Metadata Enablement Support

The following steps are imperative when enabling metadata in an enterprise.



Let us explore the following areas:

1. Metadata Evaluation

Typically, a metadata onboarding questionnaire is the first step as it will help in determining how data is defined, organized, governed, and embedded into an enterprise's operation. It lays foundation for how the Data Governance team will define, plan, measure and execute a successful and mature metadata roadmap.

The questionnaire is structured to address the 6 main areas:

1. **Ownership**: The section will help us identify key stakeholders - who owns the data, who designed the data model, the technical steward who is responsible for providing information about the database, the business steward who will be working on improving the metadata quality.

2. **Business Glossary:** This section will help us identify if the business unit has terms documented which provide a common understanding of the data supporting the business.

3. **Technical Metadata – Data Source**: This section will help us identify the different data sources owned and consumed by the team along with the data sensitivity.

4. **Technical Metadata – Data Model**: This section will help us understand the detail of how the business works to define the data that drives it.

5. **Technical Metadata – Reports**: This section will help us identify the database the reports connect to drive business decisions.

6. **Technical Metadata – Data Lineage**: This section will help us understand the end-to-end data flow.

Purpose of the Questionnaire

The questionnaire will help assess a business unit's current metadata capabilities, identify gaps, and needs for improvement. The Metadata Management team will then evaluate the current state of metadata and build a roadmap by collaborating with the business steward to address the gaps and improve the quality of enterprise metadata. It is important for our program to compare the optimized state to current state to identify the functional gaps and demonstrate how the gap will be closed.

This will help business understand their data, their data ownership, the different areas where the data is stored and the various business units consuming the same data. Most importantly, it will help identify the data flow from source to the target database from where reports and dashboards are created. If the technical metadata and business glossary don't adhere to the naming standards or definition guidelines and have inconsistent definitions it will result in difficulty to locate and understand the data. It will also be difficult for data analysts and scientists to trust the data to build dashboards.

Role of Data Governance (DG) Team

The DG team will help the business organize their data, support data sharing by ensuring compliant use of data and reduce data redundancy. In order to achieve this optimized state, we need to start by realizing the current state of metadata and gradually improve its quality by incorporating additional capabilities.

***Deliverables*:**

1. A prioritized list of data assets accompanied by a listing of corresponding metadata records that meets business needs.
2. A gap analysis report that describes the status of each dataset, its associated metadata record.

2. Recommend Scope and Process

The objective of this phase we will determine the scope, reusability of existing documentation, data model and data dictionaries and determine appropriate metadata process (including automation services) to improve the quality of metadata.

The recommended scope and process is grouped into three phases, required, recommended and optional. Implementing these metadata component takes time and resources, it is therefore not only critical to follow all the guidelines, but also to prioritize the components. It is a collaborative effort from the business steward, DG team and the DevOps.

The metadata components are summarized in the table below:

Phase	**Component**
Required	• All critical data sources identified • Business glossaries supporting the management and definition of data • Data Lineage for the critical data sources and/or attributes from source to target • Data Classification to classify sensitive data
Recommended	• Data Model designed for the source and target data source • Data Lineage from model to data sources to reports • Semantic Mapping/Classification • Report Metadata
Optional	• Taxonomies/hierarchies • Business data models • Optimizations and Automations

Successful component discovery needs to account for both business and technical focus:

Business focus:

Mandatory: **Business Glossary, Data Source, Data Model**
Optional: **Reports, Data Lineage**

Business metadata supports the business user's perspective of the data sources, by using common business terms and providing information about the source metadata in terms of context and understandability, rather than in terms of database technology.

Starting at the prioritized target system and tracing back to the source system. The target system will be linked to a business term, by this approach it will enable business users to understand where a specific data set resides. This approach will enable business users to answer question about discovering data elements across the enterprise, like "Where is customer information (specific - customer credit card, social security).

This will help harvest source metadata, data lineage, data model along with the business glossary.

Technical focus:

Mandatory: **Data Source, Data Lineage, Data Model**
Optional: **Reports, Business Glossary**

Starting at the prioritized source systems and identifying all the immediate consumers and the subsequent consumers till we reach the target database. This will help us identify the data flow for a particular source system. This approach will enable technical users to perform impact analysis like "What systems are impacted if the width of a specific column is changed?"

If there exists a data model for the target system, it will be stitched/linked to the target system.

Prioritizing the Metadata Component

Data sources that are important to prioritize are identified and designated as critical data sources during the metadata evaluation phase. This will help identify the technical metadata and business glossaries that are relevant and needed to its support the harvest. After the priorities are set, the next step will involve identifying the effort required to harvest the metadata and implement the metadata process.

For example, when there is a request to harvest metadata using bridges/connectors which hasn't been tested before, the effort taken to complete the harvest will be longer compared to bridges which have been tested.

Metadata Process

After prioritizing the metadata, the next step will be to identify the process which will enable the harvested metadata follow standards, guidelines and help automate data cataloging process. Some of the process include semantic mapping, bootstrapping, naming standards via machine learning. This process will help us achieve:

1. Clear Ownership for Metadata Management.
2. Standardized terminology which will in turn help standardize communications between stakeholders.
3. Centralization of metadata for related data.
4. Maintain a Single Source of Truth.

3. Enablement Roadmap

The enablement roadmap will provide a graduated path for successive improvements to move from the existing to the future managed Metadata environment. With each successive level, additional capabilities are achieved. Eventually it will help establish the

ownership and operational structure needed to ensure that corporate data is managed as a critical asset and implemented in an effective and sustainable manner.

4. Report and Audit Metadata

While implementing any enterprise-wide program, it is imperative that each completed phase is thoroughly reviewed before proceeding to the next phase. From that perspective, the Report and Audit Metadata phase is a standardized control/check point wherein the business steward and the DGE team meet to review and/or audit and approve the completed phase to continue with the next phase. The 3 stages are:

- **Initiated Tollgate**: State where the team discusses the customized roadmap before a domain goes to the initiation phase.

- **Performed Tollgate**: State when most of the metadata activities have been completed and the team is fully operational but need monitoring.

- **Managed Tollgate**: State when Data Governance Enablement team determines the domain is fully operational and mature and no longer needs to be monitored.

Tollgate	Average Timeline	Activities
Initiated	2-3 weeks from the time the stewardship team submits the questionnaire to the metadata management team	1. Completion of onboarding questionnaire - Business Steward 2. Review the questionnaire - DGE 3. Metadata State Evaluation - DGE 4. Follow up discussion - DGE & Business Steward
Performed	7-8 weeks or 2 months	1. Technical & Business Metadata harvesting 2. Metadata certification 3. Data Models are stitched to Data sources 4. Operation metadata harvesting
Managed	4 to 5 months	1. Monitoring and Alerting process exists 2. Metrics are defined and created for assessment.

Conclusion

While there are many benefits of implementing an enterprise-wide metadata management, the most vital outcome is data that is well defined, qualitiative, consistent and trusted. This data may be used by senior leadership can rely upon to make strategic decisions. That, this will lead to improvement of financial performance (and) optimal utilization of resources for overall operational efficiency.

However, the abovesaid benefits won't accrue instantenously. The stakeholders must be cognizant that benefits of implementing such programs are not visible immediately and may take a few years. It is imperative that progress is measured on a periodic basis and tracked for obtaining a better understanding of a return on investment.

References

https://www.datamation.com/big-data/metadata-management-tools/

https://www.dataversity.net/fundamentals-metadata-management/

https://www.talend.com/blog/2018/09/07/metadata-management-101-what-why-and-how/