

Literature Review on Metadata Governance

Amarmeet Kaur¹, Medzhakhed Usama², Nur Liana³
Doris Hooi-Ten Wong⁴, Nurazean Maarop⁵

*Razak Faculty of Technology and Informatics
Universiti Teknologi Malaysia,
54100 Kuala Lumpur, Malaysia*

*amarmeetkaur@graduate.utm.my¹,
medzhakhedusama@graduate.utm.my²,
nurlianaabmajid@graduate.utm.my³
doriswong@utm.my⁴, nurazean.kl@utm.my⁵*

Article history

Received:
25 April 2023

Received in revised
form:
18 May 2023

Accepted:
11 June 2023

Published online:
27 June 2023

*Corresponding
author
doriswong@utm.my

Abstract

The framework of metadata governance is a subset of the primary data governance framework implementation within an enterprise. Metadata management helps identify data provenance and destination systems, explain the categories of data in the system, and assist organisations in comprehending the elucidation of data in various structures. This study aims to construct a literature review on metadata governance including the overview, the definition and proposed work based on previous study in different industries related with metadata governance. According to DAMA-DMBOKv2, five activities are required to establish metadata management which are designing metadata strategy, understanding metadata needs, defining metadata architecture, producing and managing metadata, and querying, reporting, and analysing metadata. Three planning procedures for metadata management includes developing metadata strategy activities, identifying metadata needs, and designing metadata architecture. The metadata management industry would benefit from a comprehensive and effective data stewardship framework to maximise company economic value. An enterprise requires a metadata governance framework, which includes an assessment of metadata responsibilities, life cycles, and statistics, as well as how various business activities incorporate metadata. A metadata strategy guarantees that an organisation's whole data ecology is consistent.

Keywords: *metadata governance, data governance, metadata management, DAMA-DMBOK, metadata strategy, metadata architecture*

1. Introduction

Metadata is information that is produced, kept, and distributed to describe objects, allowing the needed interaction to expand existing knowledge based on enterprise requirements [1]. Organisations can use metadata to determine where data is generated, kept, and accessed [2], which reflects the specialty of metadata of recognising and locating material relevant to the user [3]. An extension of metadata, being the implementation of metadata management can further boost the efficiency of data storage processes as well as to facilitate the process for retrieving, transferring, and exchanging the data between different systems.

Metadata management helps identify data provenance and destination systems, explain the categories of data in the system, and assist organisations in comprehending the elucidation of data in various structures [4]. The framework of metadata governance is a subset of the primary data governance framework implementation within an enterprise [5]. A good metadata management procedure ensures correct metadata extraction and access to the collected metadata. All activities conducted in the context of monitoring clean metadata will give metadata characteristics of high calibre which would be capable of accommodating all cultural and property qualities derived from information on cultural artefacts generated.

There are ten knowledge areas in data governance according to the Data Management Body of Knowledge framework (DAMA-DMBOK) [2] and metadata management being one of the knowledge areas, is a key action in the management of data. Thus, this paper aims to construct a literature review on metadata governance. The scope of this study includes the overview, the definition and proposed work based on previous study in different industries related with metadata governance.

2. Overview

2.1. Types and Categories of Metadata Management

Metadata represents an organisation's understanding of its data assets and offers data information in terms of identity, description, and linkage. According to [6], metadata offers context, structure, and categorization, allowing for optimal data utilisation, retrieval, and management. A metadata repository is a collection of data asset information that is often created and updated in manageable stages throughout time. Metadata management, which includes data lineage, or the pedigree of data [7], is significant because it specifies data availability to users and enables visibility on the location and authenticity of data. It also offers transparency across hybrid and multi-cloud architectures by offering an overview of data across on-premises and cloud data repositories and employing automation to minimise the need for time-consuming physical dataset curation.

According to [6], metadata is commonly assorted into three predominant sorts:

a. Business Metadata [8]

Business metadata is applied to datasets in terms of descriptions within context and usage, owners and referents, tags, and properties to construct a taxonomic classification over the datasets that will be indexed by the search engine. Business metadata may also be found at the schema level of a dataset in the form of descriptions, tags, or the level of data confidentiality per column. A component of business metadata is the beginning point for mapping to associated permanent work items such as rules and requirements.

b. Operational Metadata

Operational metadata provides a basis for understanding when and how the data was generated or converted through statistical analysis of the data, date of update, origin (lineage), volume, cardinality, the identification of the processes that altered the data, and the status of the processes on the data, to mention a few. It may also relate to governance bodies, including their power, participation, structure, and tasks in the executive data council, for instance [9]. Additionally, operational metadata is utilised to identify process modifications in order to augment productiveness and data quality.

Process metadata [10], a subset of operational metadata, addresses process stages for data creation and maintenance, as well as data quality analysis and evaluation, such as quality standards and control needs.

c. Technical Metadata

Technical metadata describes the structure of a dataset and the information related to storage systems, in terms of changes to data sources, content, and location of data stores and interfaces. It also constitutes data type information, links to related files, and database indexes. Technical metadata is divided into the following categories [11]: 1) "run-time" or dynamic metadata, for example, configuration or message data, and 2) "design-time" or static metadata.

2.2. Metadata Management based on Data Management Body of Knowledge (DAMA-DMBOK) [2, 12]

Metadata management will aid in the organisation's processing, upkeep, consolidation, administration, monitoring, and distribution of metadata. Furthermore, metadata will assist to monitor data quality, boost data trust, and unravel data redundancies [2]. Data and data dictionary redundancy, inconsistent delineations of data items, and data abuse will ensue from incorrect metadata. Metadata management will assist organisations in understanding data, exchanging data between systems, accessing data, and sharing data [2].

In establishing metadata management, DAMA-DMBOKv2 allows for five activities: designing metadata strategy, understanding metadata needs, defining metadata architecture, producing and managing metadata, and querying, reporting, and analysing metadata. DAMA-DMBOK describes three planning procedures for metadata management: developing metadata strategy activities, identifying metadata needs, and designing metadata architecture. A control process is to query, report, and analyse information, whereas an operational process is to produce and manage metadata. The analysis is required for the design of metadata management operations.

3. Synthesis of Findings

3.1. Processes encompassing the Metadata Management based on the Data Management Body of Knowledge (DAMA-DMBOK)

Metadata management comprises a variety of activities across several industry verticals, all of which must be appropriately supported by operational procedures in order to build and maintain an appropriate and structured metadata governance framework. According to the guidelines authored in DAMA-DMBOK for metadata management, the following processes are segmented as follows:

a. Definition of Metadata Strategies [12]

The first step in developing the metadata strategic plan is to describe the methodologies for successfully producing, integrating, and applying metadata. Before considering or employing information management systems, the metadata strategy is created. A company's metadata strategy must be in line with its business aims. The process of defining metadata defines how a company wants to manage its information and improve its procedures. The metadata strategy should offer the development team with a framework for improving metadata management. Metadata development is the process of explaining the strategy and identifying potential hurdles to its implementation. The goal of establishing metadata strategies is to make business orientation the foundation for organised data sharing, to offer acknowledgment of the value of data and data constituents, to generate maps of organisational information requirements, to emphasise the relevance of data administration, data quality, data reuse, and to quantify the usefulness of knowledge.

b. Understanding of Metadata Governance Requirements [12]

The second planning phase, comprehending metadata needs, is carried out to determine what metadata and metadata levels are required. Metadata has a wide spectrum of content. This is because metadata is derived from data consumers' business and technological requirements.

c. Definition of Metadata Governance Architecture [12]

Building an architectural metadata design is the third phase in building metadata architecture. Metadata architecture must be able to regularly scan metadata sources and update metadata repositories. The system must support metadata modifications, requests, searches, and lookups of various user metadata categories. The five qualities of metadata architecture are integrated, scalable, robust, customised, and open. The operational phase, which includes developing and managing metadata, enhances metadata quality. Metadata must be maintained like a product, hence metadata generation must be meticulously planned. Accountability, standards, and improvement are three broad ideas in metadata management that may be used to improve metadata quality.

d. Metadata governance stakeholder [13]

In addition to the procedures in DAMA-DMBOK, metadata not only specifies sets of data but also includes the correlation between datasets, which is crucial for deciphering data lineage. One of several procedures in data governance is establishing the workforce and designating who is to be responsible, accountable, consulted, and informed about the data. Because of the necessity of domain expertise to analyse each data collection from a specific domain or system, subject matter experts or data stewards are in charge of metadata management. Several data stewards organise and manage metadata at the corporate level via data governance committees. Nonetheless, when corporations experience exponential data growth in the likes of the Internet of Things (IoT) environment whereby data is continually processed, mined, and stored in several forms, it is not viable for the instance of Small and Mid-Size Enterprises (SME) to enforce an effective metadata management. Therefore, the automatization of metadata is becoming progressively significant.

4. Proposed Works, Solutions and Applications

Failures in data management caused users to be unable to grasp the data in various systems. A Chinese study discovered that a lack of data communication and management planning in a railway information system hampered data retrieval, data transmission, and data exchange systems. A research by [4] offered a novel strategy for building a metadata management system for railway information systems and improving the efficiency of railway data administration. The upgraded system enabled metadata to be transferred and maintained across diverse platforms.

By incorporating the two metadata models into the system design, a centralised metadata management system was built. This enabled business workers to immediately access the information and data items of each data entity, as well as exchange the relationships between data entities in other systems. The metadata management system not only aided in identifying data assets, but it also became an essential auxiliary tool of railway information management [4], and it is extremely crucial to railway data governance.

Effective metadata management plays a significant role in establishing, enforcing, and maintaining the data governance in a smart factory, which has highly distributed IT architecture and relies on data quality for its success not just for compliance. [13] created a prototype that was created based on the proposed architecture for a single-factory organisation. They discussed the importance of data governance in Smart factory, strategy and best practices to automate metadata management. Centralised metadata repositories help to improve the metadata accuracy by 50% and reduce the edge layer compute load [14], which used to be a concern due to real-time metadata fetches unpredictable queries from data consumers. The typical duration of time to enforce enterprise level cleansing policies used to be more than 90 days. After implementing the new model, enterprise level data cleansing policies were enabled in less than a week. New

policy or code changes are reviewed and approved by the change approval board, which consists of data stewards, custodians, and owners.

A study by Tesfamicael et al. proposed a new system for the National Electricity Market. The system was designed according to the Clark-Wilson Integrity Model which recorded every transaction made through the telephone call. The system allows an ease for future telephone call audits that simultaneously represent the higher level of security and integrity of the metadata. The system assists the auditing process not just for the security and business requirements, but also to fulfil the legal requirements as stipulated by the Australian National Electricity Rules [15]. The metadata was used for the verification of the integrity of the telephone call recordings stored in the blockchain. Meanwhile, the actual telephone call recordings are stored in a separate data store. The author proposed that the design can be adopted for other commercial blockchain based applications.

All the studies above describe and propose the framework or models which were constructed and enhanced to improve metadata governance in different domains or organisations. Every organisation has its own IT landscape and different goals. These differences require customisations within its data architecture and data governance to produce a better working system for data governance particularly metadata governance.

5. Conclusion and Future Work

To recapitulate, metadata management adds value to businesses through boosting creativity, cooperation, and risk mitigation. Metadata management is used in data governance to enforce a quality system on data gathering and control. Metadata management is a vital constituent of any successful data governance approach, and metadata is one of the core contributors to an organisation's data producing and retaining full commercial value.

A metadata strategy, metadata integration and publishing, metadata capture and storage, and metadata governance and administration are all required components of a robust metadata management implementation. A metadata strategy guarantees that an organisation's whole data ecology is consistent. It also explains the rationale of commercially tracking metadata and describes all of the metadata sources and procedures that are employed. During metadata collecting, it is critical to identify all internal and external sources of metadata, which may be accomplished through the use of solutions such as metadata repositories, data modelling, and data governance tools. Ultimately, an enterprise requires a metadata governance framework, which includes an assessment of metadata responsibilities, life cycles, and statistics, as well as how various business activities incorporate metadata.

Future expectations of metadata governance should include specificity for different industries, to ease in the development of quality data with clarity and to allow enterprises to acknowledge subsets of data from different systems. Conversely, a lack of knowledge about the prospective advantages of enterprise

metadata management solutions, as well as technological problems in metadata storage and data cross-linking, are recognised as possible barriers to the worldwide enterprise metadata management market's growth. As a result, the metadata management industry would benefit from a comprehensive and effective data stewardship framework to maximise company economic value through metadata while also assisting decision-making by emphasising business needs. This would entail implementing data governance procedures, which would provide actual data consumers with value and context for understanding the data and its components.

Acknowledgments

A sincere appreciation and thanks to Dr. Doris and UTM for bearing the torch of enlightenment in providing sufficient information to aid the completion of this review paper.

References

- [1] National Information Standards Organisation. (2004). *Understanding Metadata*. USA: NISO Press.
- [2] DAMA International. (2017). *DAMA-DMBOK Data Management Body of Knowledge Second Edition*. New Jersey: Technics Publications.
- [3] Basuki, & Sulisty. (2000). Metadata, Deskripsi Serta Titik Aksesnya dan INDOMARC. *BACA*, Vol.25, No.1-2, 1-19.
- [4] Qin, D., Huang, L., & Wang, Y. (2018). Construction Of Railway Metadata Management System Based on Metadata Content Model and CMW Exchange Mechanism. 978-1-5386-6965-5/18.
- [5] Baba, A., Kaarthigaa, K., Hooi-Ten Wong, D., Zainuddin, N.M.M. and Yaacob, S., 2021. Metadata Governance Implementation Literature Review and Propose Work on Telecommunication Industry. *Open International Journal of Informatics*, 9(2), pp.1-6.
- [6] Sawadogo P, Darmont J. On data lake architectures and metadata management. *Journal of Intelligent Information Systems*. 2021 Feb;56(1):97-120.
- [7] Eichler RK. *Metadata management in the data lake architecture* (Master's thesis).
- [8] Diamantini C, Giudice PL, Musarella L, Potena D, Storti E, Ursino D. A new metadata model to uniformly handle heterogeneous data lake sources. In *European Conference on Advances in Databases and Information Systems 2018 Sep 2* (pp. 165-177). Springer, Cham.
- [9] Stathias V, Koletti A, Vidović D, Cooper DJ, Jagodnik KM, Terryn R, Forlin M, Chung C, Torre D, Ayad N, Medvedovic M. Sustainable data and metadata management at the BD2K-LINCS Data Coordination and Integration Center. *Scientific data*. 2018 Jun 19;5(1):1-4.
- [10] Eryurek E, Gilad U, Lakshmanan V, Kibunguchy-Grant A, Ashdown J. *Data Governance: The Definitive Guide*. " O'Reilly Media, Inc."; 2021 Mar 8.
- [11] Yu H, Cai H, Liu Z, Xu B, Jiang L. An Automated Metadata Generation Method for Data Lake of Industrial WoT Applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2021 Oct 21.
- [12] Khairunisak, T. F. Kusumasari and R. Fauzi, "Design Guidelines and Process of Metadata Management Based on Data Management Body of Knowledge," *2021 7th International Conference on Information Management (ICIM)*, 2021, pp. 87-91, doi: 10.1109/ICIM52229.2021.9417156.
- [13] Sundarraj, M. and Rajkamal, M.N., 2019. Data governance in a smart factory: Effective metadata management. *Int. J. Adv. Res. Ideas Innov. Technol*, 5, pp.798-804.
- [14] Oukhouya L, Er-raha B, Asri H. A generic metadata management model for heterogeneous sources in a data warehouse. In *E3S Web of Conferences 2021* (Vol. 297, p. 01069). EDP Sciences.
- [15] Tesfamicael, A.D., Liu, V., Mckague, M. and Caelli, W., 2021. An Architecture to Facilitate Security Assurance and Legal Compliance for Call Auditing in the Wholesale Electricity Market. *IEEE Access*, 9, pp.146437-146453.