ISSN: 2959-6386 (Online), Vol. 2, Issue 1

Journal of Knowledge Learning and Science Technology

journal homepage: https://jklst.org/index.php/home

Reference Data Management: A Cornerstone of Financial Data Integrity

Manish Tomar[1], Jawaharbabu Jeyaraman[2]

[1]Citibank, USA.

[2]TransUnion, USA

Abstract

Reference data management plays a crucial role in ensuring the integrity and accuracy of financial data in the banking and finance sector. This paper examines the significance of reference data management as a cornerstone of maintaining data integrity within financial institutions. It discusses the challenges faced by organizations in managing reference data effectively and explores strategies for implementing robust reference data management frameworks. By addressing these challenges and implementing best practices in reference data management, financial institutions can enhance data accuracy, improve regulatory compliance, and mitigate operational risks.

Keywords: Reference data, Data integrity, Financial data, Reference data management, Banking.

Article Information:

Article history: 13/09/2023 *Accepted:* 15/09/2023 *Online:* 30/09/2023 *Published*: 30/09/2023

DOI: https://doi.org/10.60087/jklst.vol2.n1.P144

Correspondence author: Manish Tomar

Introduction

In the fast-paced and highly regulated world of finance, ensuring the accuracy and integrity of data is paramount. Reference data management (RDM) stands as a cornerstone in the architecture of financial systems, providing the essential framework for maintaining consistency and reliability across vast datasets. Financial institutions rely on reference data to categorize and classify various financial instruments, customers, counterparties, and other entities essential for their operations.

As the volume and complexity of financial data continue to grow exponentially, so do the challenges associated with managing reference data effectively. Inaccurate or incomplete reference data can lead to erroneous reporting, regulatory non-compliance, and increased operational risks. Therefore, it is imperative for financial organizations to develop robust reference data management strategies to ensure data integrity and regulatory adherence.

This paper delves into the significance of reference data management in the financial sector, highlighting its role in maintaining data integrity and supporting critical functions such as risk management, trading, and regulatory reporting. It explores the complexities and challenges inherent in managing reference data, including data governance, standardization, and interoperability across systems and platforms. Additionally, the paper discusses emerging trends and best practices in reference data management, aiming to provide insights into how financial institutions can enhance their data management processes to meet regulatory requirements and improve operational efficiency.

Literature Review:

Data management is a crucial aspect of ensuring financial data integrity [1]. It plays a significant role in research data integrity, which is essential for research rigor, reproducibility, and data reuse [2]. Data management involves planning and implementing actions throughout the research data lifecycle, such as data acquisition, analysis, and preservation [3]. It is closely associated with data quality and data security, as reliable, trustworthy, valid, and secure data are necessary for maintaining data integrity [4]. In addition, data management is gaining momentum in solving challenges related to data ownership [5]. Proper data management practices can help prevent unauthorized access, modifications, and manipulations, reducing the risk of financial and reputational losses . Therefore, data management is indeed a cornerstone of financial data integrity, ensuring accurate and secure processing of financial data .

Research Approach

A reference model serves as a benchmark in the design of an information system, representing a class of use cases and providing a framework for developing customized models tailored to specific domains (Schu¨tte 1998, pp. 69–74). Reference models vary based on factors such as application domain, modeling language, size, design process, and evaluation strategy (Fettke and Loos 2004).

This paper outlines the design process and resulting reference model for Master Data Quality Management (MDQM) functionality. Following the ARIS concept, which delineates between four descriptive views (functional, data, control, and organizational) and three descriptive layers (business design, technical design, and implementation layer), the functional reference model for MDQM represents the business design of MDQM systems, focusing on purpose and tasks rather than technical details (Scheer 1992, 1997; Scheer et al. 2005).

Aligned with the guidelines for Design Science Research (DSR) proposed by Hevner et al. (2004), the design process follows the principles of the Design Science Research Methodology (DSRM) (Peffers et al. 2008). This methodology advocates for a sequential design process involving iterations of design and evaluation cycles, with flexibility in approach. The research process begins with a problem-centered initiation, identified through focus groups rather than preconceived deficiencies in existing artifacts.

The research context is shaped by the Competence Center Corporate Data Quality (CC CDQ) at the Institute of Information Management, University of St. Gallen. Since 2006, researchers at the institute, in collaboration with partner companies, have been developing design artifacts in corporate data quality management.

Research Process

Following the DSRM process model, the design of the reference model progresses through six steps.

The initial step, conducted between January and December 2008, involves identifying the problem and motivating the research. The impetus for this research stems from challenges identified within the practitioner community. In 2008, the MDQM market witnessed significant consolidation, prompting practitioners within the CC CDQ to express a need for support in addressing various challenges. This demand was reinforced by managerial publications, such as Gartner's documentation of common queries on data integration and quality from their 2008 MDM summit, reflecting the industry's concerns and interests (Friedman 2009).

Related Work

Data Quality Management

The field of data quality management has been extensively studied, with research efforts yielding various insights and approaches. Some studies, such as those by Wang and Strong (1996), focus on identifying dimensions of data quality through empirical research, while others, like the works of English (1999), Loshin (2001), and Redman (1996), provide valuable insights from practitioners' experiences. Theoretical perspectives on data quality are also explored in studies by Price and Shanks (2005) and Wand and Wang (1996). Despite the diversity in approaches, there is a consensus that data quality is contingent on its fitness for specific use cases and user contexts.

Data quality management (DQM) encompasses efforts aimed at enhancing the quality of data (Batini and Scannapieco 2006). Unlike reactive approaches focused solely on identifying and rectifying data defects, DQM adopts a proactive stance. It employs iterative processes involving steps such as defining, measuring, analyzing, and improving data quality, alongside designing suitable frameworks for DQM (English 1999; Wang 1998; Eppler and Helfert 2004). Batini et al. (2009) provide an overview of various DQM methodologies and approaches.

Master Data Management

Master data constitutes the core business entities upon which an organization's activities rely. These entities encompass crucial aspects such as business partners (customers, suppliers), products, and employees (Smith and McKeen 2008). Conceptually, master data can be categorized into master data classes, attributes, and objects (Loshin 2008). A master data object represents a specific business entity (e.g., an automobile produced at a particular plant at a specific time) and is characterized by attributes defining its properties (e.g., color, features, price).

Master data management (MDM) encompasses the entirety of activities involved in creating, modifying, or deleting master data classes, attributes, or objects (Smith and McKeen 2008; White et al. 2006). This includes tasks such as modeling, provisioning, quality management, maintenance, and archiving of master data. The overarching goal of MDM is to ensure the availability of high-quality master data—data that is complete, accurate, timely, and well-structured—to support various business processes (Loshin 2008; Karel 2006).

Integration of MDM and DQM

The recognition of data quality as a fundamental objective within Master Data Management (MDM) has led to the consideration of Data Quality Management (DQM) as one facet of MDM in several studies (DAMA 2009; Dreibelbis et al. 2008). However, this perspective often confines the analysis of DQM to reactive measures only (White and Radcliffe 2008). Upon closer examination, it becomes evident that there are intertwined activities in both MDM and DQM domains, each influencing the other. Thus, any attempt to subordinate one area to the other proves inadequate.

In particular, proactive DQM initiatives such as data governance (Weber et al. 2009; Khatri and Brown 2010) or business metadata management (Burnett et al. 1999; Marco 2000) play a significant role in shaping an organization's MDM framework. These preventive DQM activities contribute to the structuring of MDM by defining master data elements and delineating responsibilities for data maintenance. Consequently, preventive DQM influences the design and implementation of MDM within an organization.

Given this interplay between MDM and DQM, the reference model proposed in this paper does not seek to prioritize one concept over the other. Instead, it delineates business requirements for application systems that support both MDM and DQM activities. The overarching goal is to ensure the provision of high-quality master data to facilitate efficient business processes.

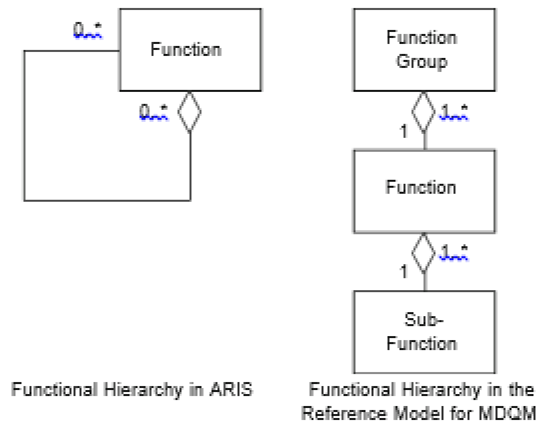
Reference Model Design

Design Foundations

The design of the functional reference model for Master Data Quality Management (MDQM) adheres to the ARIS conventions for the functional view of information systems (Scheer 2001, pp. 21–38), which advocate for a hierarchical structure of functions. The reference model described in this paper comprises three levels: function groups, functions, and sub-functions (refer to Fig. 2).

Functions within the MDQM reference model are organized into function groups. Each function group contains one or more functions, with each function belonging to only one function group. Functions themselves are composed of sub-functions. The relationships between functions and sub-functions mirror those between function groups and functions. This three-tiered hierarchical structure aligns with the modeling principles established in ARIS (Scheer 2001, p. 25).

Fig. 2 Modeling functional hierarchies



In essence, functional hierarchies are typically formed based on three criteria: performance, object, and process. The reference model for Master Data Quality Management (MDQM) adopts the process criterion, organizing functions into function groups and sub-functions into functions based on their purpose-oriented and task-oriented relationships.

The visual depiction of the reference model adheres to the principles of process maps, which aim to identify and represent similar processes, sub-processes, and functions in a tabular format (Heinrich et al., 2009). This approach is commonly utilized in the practitioners' community, exemplified by SAP's business maps. Specifically, technology-related business maps, such as those for SAP NetWeaver, employ a tabular design (SAP, 2007a). The decision to utilize a tabular presentation format in this paper was deliberate, chosen to ensure high comprehensibility and acceptance among potential users of the reference model.

Model Overview

The reference model encompasses a total of 6 function groups, 19 functions, and 72 sub-functions. Figure 3 illustrates the function group and function levels of the model (with all 72 sub-functions detailed in the Appendix).

Below is a description of the six function groups:

- Master Data Lifecycle Management: This group entails activities associated with the entire lifespan of a master data object, from its creation during business operations to its deactivation or archiving. Functions within this group, such as Create or Update, are inherently self-explanatory.

- Metadata Management and Master Data Modeling: Metadata defines data properties and meanings, including those of master data. It plays a crucial role in specifying data structures and ensuring correct data usage throughout an organization.

Master Data Lifecycle Management	Data Creation	Data Maintenance	Data Deactivation	Data Archiving
Metadata Management and Master Data Modeling	Data Modeling	Model Analysis	Metadata Management	
Data Quality Assurance	Data Analysis	Data Enrichment	Data Cleansing	
Master Data Integration	Data Import	Data Transformation	Data Export	
Cross Functions	Automation	Reports	Search	Workflow Management
Administration	Data History Management	User Management		

The organization (Burnett et al., 1999; Marco, 2000; Tozer, 1999) delineates several key function groups within the context of Master Data Management (MDM). From the MDM standpoint, metadata encompasses all requisite information for the efficient management and effective utilization of master data. In this context, master data modeling involves the creation of technical metadata, encompassing data types and relationship multiplicities.

The function groups include:

- Data Quality Assurance: This group encompasses functions aimed at both preventive and reactive maintenance and enhancement of master data quality. These functions involve identifying data defects, measuring data quality (Data Analysis), enriching data through comparison and integration with external reference sources (Data Enrichment), and rectifying identified data defects (Data Cleansing).

- Master Data Integration: Functions in this group facilitate the transfer (import and export) and structural transformation (e.g., consolidation of fields or tables) of master data.

- Cross Functions: This group includes functions that cannot be categorized under other groups. Sub-functions within the Automation function do not introduce additional functionality but provide support for enabling efficient utilization of other functions by making them machine-processable.

- Administration: This group comprises functions related to user administration and tracking changes and modifications made within the system.

Conclusion

The paper outlines a functional reference model for Master Data Quality Management (MDQM), developed through a rigorous design process following the principles of the Design Science Research Methodology (DSRM). The model, comprised of six function groups and spanning 19 functions and 72 sub-functions, serves as a valuable tool for both advancing scientific knowledge and enhancing practical applications in the field.

Practitioners can leverage the reference model to analyze and design MDQM systems within their organizations, facilitating software evaluation and fostering communication both within and across companies. From a research standpoint, the model represents an information system, elucidating business user requirements and contributing new insights into real-world applications.

However, the reference model has its limitations, primarily focusing on the business layer of MDQM and excluding other aspects such as control and organizational views. Consequently, its application is restricted to similar use cases. Future research should aim to expand the model by incorporating additional views and levels, particularly exploring the control and organizational perspectives. Through case studies, researchers can identify generic characteristics of DQM and MDM organizations, informing the conceptualization of rights and roles within the organizational view. Additionally, further investigation into the interdependencies between functions and typical MDQM activities could enhance the model's comprehensiveness and applicability.

References List:

1. Alur, N., Joseph, R., Mehta, H., Nielsen, J.T., Vasconcelos, D. (2007) IBM WebSphere Information Analyzer and Data Quality Assessment. IBM, Armonk

2. Baskerville, R.L. (1997) Distinguishing Action Research from Participative Case Studies. Journal of Systems and Information Technology 1(1):24–43. doi:10.1108/13287269780000733

3. Batini, C., Scannapieco, M. (2006) Data Quality: Concepts, Methodologies and Techniques. Springer, Berlin.

4. Batini, C., Cappiello, C., Francalanci, C., Maurino, A. (2009) Methodologies for Data Quality Assessment and Improvement. ACM Computing Surveys 41(3):1–52. doi:10.1145/1541880.1541883

5. BEA (2006) BEA AquaLogicTM Data Services Platform. BEA Systems, Inc., San Jose.

6. Becker, J., Rosemann, M., Schütte, R. (1995) Grundsa¨tze ordnungsma¨ßiger Modellierung (Guidelines for Orderly Modeling). Wirtschaftsinformatik 37(5):435–445.

7. Becker, J., Algermissen, L., Delfmann, P., Knackstedt, R. (2002) Referenzmodellierung (Reference Modeling). Das Wirtschaftsstudium (WISU) 30(11):1392–1395.

8. Becker, J., Holten, R., Knackstedt, R., Niehaves, B. (2004) Epistemologische Positionierungen in der Wirtschaftsinformatik am Beispiel einer konsensorientierten Informationsmodellierung (Epistemological Positionings in IS Using the Example of a Consensus-Oriented Information Modeling). In: Frank, U. (ed) Wissenschaftstheorie in O¨konomie und Wirtschaftsinformatik (The Theory of Science in Economics and IS). Deutscher Universita¨ts-Verlag, Wiesbaden, pp 335–366.

9. Beiersdorf (2009) Annual Report 2009. Beiersdorf AG, Hamburg.

10. Burnett, K., Ng, K.B., Park, S. (1999) A Comparison of the Two Traditions of Metadata Development. Journal of the American Society for Information Science 50(13):1209–1217. doi:10.1002/(SICI)1097-4571(1999)50:13<1209:AID-ASI6>3.3.CO;2-P.

11. Butler, D. (2007) MDM as a Foundation for SOA. Oracle Corporation, Redwood Shores.

12. Deloitte (2009) Telecommunications Predictions—TMT Trends 2009. Deloitte Touche Tohmatsu, London.

13. Dreibelbis, A., Hechler, E., Milman, I., Oberhofer, M., van Run, P., Wolfson, D. (2008) Enterprise Master Data Management: An SOA Approach to Managing Core Information. Pearson Education, Boston.

14. English, L.P. (1999) Improving Data Warehouse and Business Information Quality, 1st edn. Wiley, New York.

15. Eppler, M.J., Helfert, M. (2004) A Classification and Analysis of Data Quality Costs. In: Chengalur-Smith, S., Long, J., Raschid, L., Seko, C. (eds) Proceedings of the 9th International Conference on Information Quality, Cambridge, pp 311–325.

16. Fettke, P., Loos, P. (2004) Referenzmodellierungsforschung (Reference Modeling Research). Wirtschaftsinformatik 46(5):331–340.

17. Vemuri, N. V. N. (2023). Enhancing Human-Robot Collaboration in Industry 4.0 with AI-driven HRI. *Power System Technology*, *47*(4), 341-358. Doi: https://doi.org/10.52783/pst.196

18. Vemuri, N., Thaneeru, N., & Tatikonda, V. M. (2023). Smart Farming Revolution: Harnessing IoT for Enhanced Agricultural Yield and Sustainability. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(2), 143-148.
DOI: https://doi.org/10.60087/jklst.vol2.n2.p148