Improving Data Quality and Data Governance Using Master Data Management: A Review

Abstract—Master data management (MDM) is a method of maintaining, integrating, and harmonizing master data to ensure consistent system information. The primary function of MDM is to control master data to keep it consistent, accurate, current, relevant, and contextual to meet different business needs across applications and divisions. MDM also affects data governance, which is related to establishing organizational actors' roles, functions, and responsibilities in maintaining data quality. Poor management of master data can lead to inaccurate and incomplete data, leading to lousy stakeholder decision-making. This article is a literature review that aims to determine how MDM improves the data quality and data governance and assess the success of MDM implementation. The review results show that MDM can overcome data quality problems through the MDM process caused by data originating from various scattered sources. MDM encourages organizations to improve data management by adjusting the roles and responsibilities of business actors and information technology (IT) staff documented through data governance. Assessment of the success of MDM implementation can be carried out by organizations to improve data quality and data governance by following the existing framework.

Keywords—MDM, Master Data, Data Quality, Data Governance.

I. INTRODUCTION

Along with the development of information technology, reliable and trusted information system governance is necessary for every organization. Information systems must be trustworthy and dependable to ensure data quality, data consistency, data accuracy to support business decisions and become the foundation for future organizational growth [1].

Data is one of valuable assets owned by an organization. One of the problems faced by many organizations is that data is managed individually by various functional and structural units. There are differences in the data format used by each part of the organization, which results in data uniformity [2]. If the data is managed correctly, it will produce information that the organization can use for decision-making.

Increasing the amount of data is a challenge for organizational data management. It causes data quality issues that are very common in organizations today. Data quality remains a big problem for organizations which information systems spread over several different units or departments.

Such information systems allow for the resolution of data quality problems such as duplication and inconsistency. Using

ensure good data governance, the organization must determine the function of the data owner and the role of data steward [9].

B. Master Data Management

MDM helps organizations standardize the definitions and attributes of data elements (customer, vendor, product, etc.). It can facilitate data sharing among all business functions, departments, and even different divisions across all information systems, platforms, and applications [10]. MDM creates a single data view of the targeted data domain. For example, master data management for customer records will refer to the "single truth" or "single customer view" of customer records. The process of making records by extracting, cleaning data from various corporate data sources is called customer data integration (CDI). CDI is part of MDM. In various data sources inside an organization, MDM may provide solutions to solve quality problems like duplication, inaccuracies, and incoherence.

The main process of MDM is profiling master data. The master data profiling feature aims to assess the state of data quality at various sources. Second, consolidating the master data into a repository and linking it with various existing applications. Third, cleaning the master data and enriching information. Fourth, synchronizing master data with organizational business processes with connected applications to support business intelligence and reporting systems [11].

MDM supports organizations integrating and sharing data in an accurate, timely, and consistent manner by implementing policies, services, and infrastructure. It guides the organization in creating master data. A successful MDM solution requires several things, namely data inventory used throughout the organization, the ability to share and send information, identification of data objects that are integrated with the master data, and governance framework to manage integration on an ongoing basis.

Technology is an important part of MDM [12]. However, organizational process is the most important part of MDM, not only technology and information systems [13]. The following are the steps for successful MDM process [13].

1. Identify organizations need.
2. Identify organization's master data and system that use it.
3. Define MDM governance.
4. Define process maintenance.
5. Define data standards.
6. Define future improvement for current data quality.
7. Planning MDM architecture (application, data flows, data security, and data privacy issues).
8. Training and communication for all stakeholders.
9. Roadmap/strategy for MDM development.
10. Define MDM characteristic (user interface, workflow, data editing, reconciliation, integration).

C. Data Quality

Most of the world's databases have large amounts of data, are inconsistent, and are missing due to various factors that indicate insufficient data quality. Every year, low data quality costs governments and private businesses billions of dollars in

income. Data quality problems are expected to cost companies up to 12% of their sales. Using the garbage in, garbage out (GIGO) principle for data analysis, processing, and management. If the input data quality is not good, the resulting output data will not be standard and will affect data accuracy. Input control also one of the factors that affect data quality [14].

Some researchers have defined data quality dimensions. Reference [15] has determined 15 data quality dimensions and divided them into four data quality classes. Reference [16] has provided 51 data quality dimensions classified into nine data quality classes. Reference [17] summarized six data quality dimensions which were relevant to data matching in 2012, namely accuracy, completeness, consistency, timeliness, accessibility, and believability. The UK Working Group of the Data Management Association International identified six dimensions to assess data quality in 2013. The data quality dimensions are accuracy, uniqueness, completeness, validity, timeliness, and consistency [18].

Based on [19], [20], and [21], data quality dimensions are divided into four categories, namely intrinsic, contextual, accessibility, and representational category. Poor data quality in information system has led to a problem of duplication, inaccuracy, and inconsistency of information. Issues related to data quality are increasingly important to analyze in an organization's information system. Data quality is highly dependent on various scattered data sources that influence organizational decision-making [22].

D. Data Governance

The policies and procedures used to manage data in an organization are referred to as data governance. Data governance is a process that ensures the organization works well with data assets and ensures that the organization's data assets are well maintained. Organizations benefit from data governance because of data standardization, effective business policy formulation, and stakeholders' role. Organizations may use data governance to align data management with business priorities, ensure regulatory compliance, and mitigate risk [23].

Many businesses have adopted big data in this age of digital transformation. Big data implementation necessitates a new collection of governance policies on which to base information management. Big data governance refers to an organization's accountability in information governance, stewardship, data definition, metadata management, master data management, usage standards, data life cycle management, risk, and cost control. Big data governance includes optimization, privacy, and monetization policies following the objectives set. To control, ensure the availability, useability, integrity, accuracy, audit capacity of big data, the big data management system must establish policies, processes, and standards [24].

BI. RESEARCH METHODOLOGY

The study of how MDM improves data quality and data governance is carried out through literature studies. This study focuses on how MDM improves the data quality and data governance issues and assess the success of MDM implementation. Several articles related to the topic both

No	Data Quality Barriers	References
1.	Unclear master data definition	[25], [26]
2.	Lack of responsibilities for data maintenance	[25], [27]
3.	Lack of data control routines	[14], [25], [26], [27], [28]
4.	Lack of procedures	[14], [26], [28]
5.	Lack of training data users	[14], [27], [28]
6.	Lack of rewards to ensure the validity of master data	[27], [28]
7.	Lack of understanding IT system	[28]
8.	Irresponsible employee	[14], [28]
9.	Lack of data quality measurement	[26], [28]
10.	Lack of understanding of the importance of correct data	[28]

No	Data Governance Barriers	References
1.	Unclear data ownership	[9], [13], [18], [26], [28]
2.	Unclear roles and responsibilities of organizational actors in managing master data	[9], [13], [18], [26], [27], [28],
3.	Lack of procedures and policies for data management	[25], [27], [28]
4.	Lack of support from management.	[13], [25], [26], [28]

published in national and international journals were reviewed. Other supporting literature was also used, such as books or patents. This literature review contains the author's opinions and views resulting from the research carried out.

IV. RESULT

A. Data Quality and Data Governance Barriers

Poor data quality, such as inaccuracies, redundancy, and inconsistencies, is affected by several things. Based on the results of the literature analysis, there are barriers in obtaining high-quality data for organizations. Further details of data quality barriers are shown in Table I.

In dealing with data quality, organizational management requires data governance that helps employees effectively utilize company information assets. Data quality management requires the participation, collaboration, and views of all employees. The data governance model in the organization at least consists of technical leadership, oversight, steering, and stewardship tasked with ensuring good data quality [29]. Participants in data governance work together to define a data quality strategy as well as to develop and establish data quality standards to achieve the expected data quality.

Data governance is related to the roles and responsibilities of organizational actors in managing data. These problems emphasize data ownership, roles, and responsibilities. In addition, they are challenges from data governance which aims to establish policies and procedures to maintain master data.

Challenges that often take place in data governance are shown in Table II.

Data governance issued in an organization are obstacles to maintain data quality and MDM success. Proper delegation of roles and responsibilities for data management to organizational actors is important to maintain high-quality data [30]. Reference [31] describes data governance as part of MDM and organizational behavior to improve and maintain data as a strategic organizational asset.

B. How MDM Improves Data Quality and Data Governance

Master data has an important role in an organization as it becomes a reference for data transactions and changes rarely occur. Data quality assurance must occur across the master data lifecycle for MDM implementation to be effective. Improving consistency and unclear roles as well as responsibilities to maintain data quality are significant challenges in managing master data. Therefore, it is necessary to guarantee the data quality in implementing MDM so that organizations do not face repeated risks in creating master data.

MDM defines data quality into the data management process through documented roles and responsibilities under data governance [32]. To accomplish these tasks, the organization needs to establish team roles and responsibilities. The team formed is a collaboration between business people and information technology (IT) staff [9], [13], [26]. On the other hand, successful MDM implementation depends on good data governance. The data governance program will also benefit the MDM program to strengthen the ability to manage all organization information activities. Ownerships, roles, and responsibilities in data governance are emphasized through MDM. Several studies have shown that MDM resolves data quality and data governance issues. The related research is shown in Table III. With duplication, inconsistent and inaccurate data from various data sources, organizations need an integrated data management. One of the factors driving the organization to implement MDM is the desire for consistency and accuracy organizational data.

The MDM program's core is consolidating multiple data sets representing master data objects such as customers or employees. These capabilities rely on metadata and data standards discovered through data profiling and discovery processes to parse, standardize, match, and resolve [9], [17], [18], [26], [31]. Since the MDM program is intended to create high-quality master data storage that is integrated, synchronous, and consistent, the organization can perform data profiling and standardization and identify a set of master objects to assess the data's quality. However, the organization must establish resolution and management data as well as analyze the master object relationship.

MDM is not only about technology, but also all matters needed to manage master data. These matters include people, processes, and technology. According to [33] and [34], to support the implementation of good data governance in implementing MDM, a business team and IT staff are needed. The roles and responsibilities in data governance are shown in Table IV [33], [34]. The roles and responsibilities of organizational actors to manage master data must be

Issues	References
Data quality	
• Duplication	[2], [7], [9], [17]
• Accuracy	[2], [7], [9], [17]
• Consistency	[2], [7], [9], [17]
Data governance	[7], [32], [33], [34]

Role	Description
Data owner	The data owner oversees the information presented. The data owner is also in order of managing and developing the data and making decisions about the data domain they own.
Data steward IT	The data element standard format and implementation details are given. Determine any current or ongoing data quality criteria for data.
Data steward (business), information steward	Setting data standards and policies from a business perspective and defining objectives for data quality.
Data governance council	Control the development and implementation of data governance.

documented and implemented based on the rules set by the organization. MDM enhances data quality and data management processes by recording roles and responsibilities in data governance. Furthermore, MDM ensures responsibility for the maintenance of high-quality master data [9], [30], [32].

C. MDM Implementation

Organizations need to define roles, functions, and responsibilities in data governance clearly so that data quality is maintained to support MDM implementation following organizational needs [9], [18], [33], [34].

MDM encourages the formation and change of roles as well as responsibilities of organizational actors for data management and information systems to maintain high-quality master data [35]. On the other hand, a successful MDM implementation depends on high-quality master data and good data governance [9], [18], [36].

In addition to the large amount of data scattered from various sources, several issues including unclear roles, functions, and responsibilities of personnel in the organization as well as the absence of adequate policies and procedures for data management are obstacles in making master data. Therefore, a simple delegation of tasks to manage master data affects master data quality and MDM implementation.

According to previously discussed explanation, the introduction of MDM has a significant impact on the roles and responsibilities of organizational actor for data management. Yet, this linkage received little attention. Several studies conducted has indicated the roles and responsibilities of organizational actors in supporting data quality maintenance and MDM implementation. However, these roles and responsibilities have not been managed and adapted to organizational needs.

There are several challenges in implementing MDM. These challenges are explained as follows.

1) Data Ownership: The organization does not have a clear description of the data ownership concept [2], [9], [18], [26], [28], [35], [37]. Unclear roles of data owners can result in a poor data quality since the data owners are responsible for their data and ensure that the data is correct.

2) Roles and Responsibilities: The organization does not have clear documentation regarding the roles and responsibilities of employees who are responsible for maintaining data quality, such as data stewards [9], [18], [26], [35], [38]. The undocumented roles and responsibilities result in employees not carrying out their responsibilities following the job description and hand over their responsibilities to other employees.

3) Organizational Change: These problems are related to processes and people in charge of handling master data in the organization. In practice, the organization does not clearly define the data format, the flow of the data entry process, how to update the data, or who is responsible if data-related problems arise [26], [28], [37], [38]. These problems have the potential to make data entry not uniform, redundant, and inaccurate.

D. Assessing MDM Implementation

Organizations can develop a roadmap to improve MDM implementation by assessing current conditions and considering the level of maturity needed to achieve organizational goals following organizational needs. Several studies have made models for measuring MDM implementation's maturity level through the master data management maturity model (MD3M). MD3M is a measure of an organization's ability to improve MDM sustainably. The purpose of assessing the maturity level of MDM implementation is to provide opportunities for organizations to evaluate the maturity level and benchmarks of MDM against other organizations [39]. The higher the MDM implementation value, the more optimal the organization will be in preparing and managing quality master data. Several comparisons of studies related to MD3M are described in Table V.

Reference [39] presents the MDM maturity model focusing on five areas, namely profiling data sources, definition of data strategy, definition of a data consolidation plan, maintaining data, and utilization of data. The model addresses the main focus areas of MDM. However, the structure only provided a large area of interest and was not detailed considering MDM.

Reference [40] provides an overview and shows a good level of maturity. In addition to that, it suggests awareness of data quality and consistency issues, strategy or roadmap to solve master data duplication, plan MDM initiatives, MDM management support, and integration. However, the model did not clearly state the focus areas to be evaluated to assess the maturity level of MDM implementation. It only stated the characteristics that must be met from each level to be assessed.

TABLE V
COMPARISON OF MDM MATURITY MODELS

Reference	Maturity Level	Dimension
[39]	1. Marginal 2. Stable 3. Best practice 4. Transformational	1. Profiling data sources 2. Definition of data strategy 3. Definition of a data consolidation plan 4. Data maintenance 5. Data utilization
[40]	1. Initial 2. Repeatable 3. Define process 4. Manage and measurable 5. Optimized	1. Data model 2. Data quality 3. Usage and ownership 4. Data protection 5. Maintenance
[41]	0. Ignorant 1. Initial 2. Isolated 3. Organized 4. Unified 5. Optimized	1. No MDM awareness 2. Awareness of data quality issues 3. Strategy solving duplicate master data 4. Plan MDM initiatives 5. Defined MDM management support 6. Integration

It is suitable only for the first insight into an organization's MDM [40].

Reference [41] provides the main topic and focus area of MD3M, covering all aspects of MDM important to the organizations. MD3M is divided into five main topics, namely data model, data quality, usage and ownership, data protection, and maintenance. These main topics are intended to assess the maturity of master data management. Those five main topics are divided into 13 focus areas. A data model's focus area is on data definition, data model, and the data landscape. Data quality focuses on assessments, impact on business, awareness of quality gaps, and improving data quality. The focus of use and ownership is data use, ownership, and access. Data protection focuses on data security or protection. Finally, maintenance focuses on data storage and life cycles [41], [42]. This model's implementation is customized to the organization's need associated with master data by offering a model for evaluating maturity that has been validated in research [43]. To assess the implementation of MDM maturity, the organization can assess from which dimension are the main and least mature areas of concern. Organizations can use these evaluations to improve master data management and compare organizational efficiency with other organizations.

V. CONCLUSION

Based on the review results, it can be concluded that MDM can solve data quality issues through the MDM process like data profiling and standardize, caused by data originating from various scattered sources. MDM also influences data governance. MDM encourages organizations to improve data management by adjusting the roles and responsibilities of business actors and IT staff documented through data governance. Data governance contains the roles, functions, and

responsibilities of business actors and IT staff to support MDM implementation. A clear delegation of roles, functions and responsibilities in data governance also impacts data quality because good data governance can maintain high-quality master data.

Assessment of the success of an organization in implementing MDM to improve data quality and data governance needs to be done. In addition, several steps need to be taken in the research method step. Data collection such as observation, interviews, and discussions need to be carried out to sharpen the analysis. Organizations can describe the results of the MDM implementation assessment in the form of diagrams or tables to make it easier for organizations to evaluate their MDM implementation. Those description can also be the basis for the organization to increase the level of MDM implementation following the available framework.

REFERENCES

[1] B.A. Nugraha, R.W. Witjaksono, and R. Mulyana, "Analisis dan Perancangan Master Data Management (MDM) Berbasis DAMA-DMBOK v2: Studi Kasus: PT Kereta Api Indonesia," *e-Proceeding of Art & Design*, Vol. 5, No. 3, pp. 3282-3289, Dec. 2018.

[2] R. Vilminko-Heikkinen and S. Pekkola, "Organizational Issues in Establishing Master Data Management Function," *Proceedings of the 17th International Conference on Information Quality (ICIQ 2012)*, 2012, pp. 1-13.

[3] S. Lerche, "Achieving Customer Data Integration through Master Data Management in Enterprise Information Management," Master thesis, University of Johannesburg, Johannesburg, South Africa, 2014.

[4] R.F. Smallwood, *Information Governance: Concepts, Strategies, and Best Practices*, Hoboken, USA: John Wiley & Sons, 2014.

[5] A.M.I. Purnama, "Perancangan Arsitektur Manajemen Master Data (Studi Kasus: PT.Jaya Mandiri Gemasejati)," Master thesis, Universitas Komputer Indonesia, Bandung, Indonesia, 2014.

[6] S. Wieczorek, A. Stefanescu, and I. Schieferdecker, "Test Data Provision for ERP Systems," *2008 1st International Conference on Software Testing, Verification, and Validation*, 2008, pp. 396-403.

[7] D. Cervo and M. Allen, *Master Data Management in Practice: Achieving True Customer MDM*, 1st ed., Hoboken, USA: Wiley, 2011.

[8] F. Rivard, G.A. Harb, and P. Meret, *The Transverse Information System: New Solutions for IS and Business Performance*, Hoboken, USA: Wiley-ISTE, 2009.

[9] D. Loshin, *Master Data Management*, Burlington, USA: Morgan Kaufmann, 2009.

[10] P. Kumar, Master Data Management (MDM) – Strategies, Architecture and Synchronisation Techniques, 2015.

[11] Indrajani, "Master Data Management Model in Company: Challenges and Opportunity," *ComTech: Computer, Mathematics and Engineering Applications*, Vol. 6, No. 4, pp. 514-524, Dec. 2017.

[12] A. Joshi, "MDM Governance: A Unified Team Approach," *Cutter IT Journal*, Vol. 20, No. 9. 30-35, Aug. 2007.

[13] R. Vilminko-Heikkinen and S. Pekkola, "Establishing an Organization's Master Data Management Function: A Stepwise Approach," *2013 46th Hawaii International Conference on System Sciences*, 2013, pp. 4719-4728.

[14] H. Xu, A. Koronios, and N. Brown, "Managing Data Quality in Accounting Information Systems," in *IT-Based Management: Challenges and Solutions*, L.A. Joia, Ed., Hershey, USA: IGI Global, 2002, pp. 277-299.

[15] R.Y. Wang and D.M. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems*, Vol. 12, No. 4, pp. 5-33, 1996.

[16] T.C. Redman, *Data Quality. The Field Guide*, 1st ed., Newton, USA: Digital Press, 2001.