A Framework for Improving Data Quality Throughout the MDM Implementation Process

Chafika Benkherourou[1],*Abdelhabib Bourouis[2]

[1] *University of Batna 2, Batna 05000, Algeria*
[2]*Research Laboratory on Computer Science's Complex Systems, University of Oum El Bouaghi,Oum El Bouaghi, 04000, Algeria*

*Corresponding author. Email: mailto:cbenkherourou@gmail.com

ABSTRACT

Controlling data value is becoming more challenging for businesses. Data redundancy is the most common cause of problems. Master Data Management (MDM) may help avert various issues by providing a "single version of the truth" throughout the organization. One of the fundamental causes for MDM implementation failure is the absence of participation of all stakeholders in the data quality process. This paper aims to provide a framework for improving the quality of Master Data Management project development. The framework is divided into five steps that are presented in a logical order. Our technique kicks off the data quality procedure, which is the study's main contribution. Furthermore, by using the BPMN 2.0 standard in construction processes, both business and IT workers can understand them. The paper also helps firms streamline and organize the process of building a Master Data Management project.

Keywords: Master Data Management, Data Quality, Business Process Modelling Notation

1. INTRODUCTION

Organizations are now confronted with the difficulty of managing and determining the value of their data due to the expansion of data volume. However, if the data is of low quality, it might be challenging to handle. Master Data Management (MDM), which promises to remove data duplication, decrease redundancy, and improve data quality, has recently received a lot of SI attention. As a result, implementing MDM ensures that the business does not utilize several versions of the same master data in different data sources.

On the other hand, MDM has been seen as a curative technique in recent years in the literature. This solution was implemented only on a technical level, with little regard for the business. Companies use market software products to manage their data, programs, and operations. Thus, establishing an MDM project is achieved if it is viewed only as a technology-supported phenomenon [1]. One of the critical reasons for MDM implementation failure is the lack of all stakeholders [2]. According to the literature assessment, research in this field does not address how technical problems are interlaced with business considerations to assure data quality in the MDM project's establishment. We are concentrating on the early phases of MDM installation to solve these challenges. The strategy to data quality must be set early on, and all stakeholders must be included to guarantee the implementation's success.

This article outlines a methodology for assessing data quality issues that may develop while installing a master data management system in an organization. There are five stages in the framework, which are presented in a logical sequence. It demonstrates how the data quality process is planned at an introductory level. Furthermore, adopting the BPMN 2.0 standard in creating methods ensures that both business and IT personnel can comprehend them.

To represent the data quality process, the Business Process Modelling Notation (BPMN) is employed. For this research, BPMN provides a model that business analysts and technical developers can use to help design and develop processes. The original motivation for BPMN's creation was to provide a link between business process design and execution. The second reason is that it is simple to grasp and intuitive. As a result, business analysts and technical developers should be familiar with any generally used modelling notation.

The recommended method considers the company's particular business and IT needs. The process of design

and assessment is presented in the context of a real-life scenario.

The remainder of the paper is organized as the following manner. The next section provides a summary of recent and relevant literature. The proposed framework is described in section 3. Section 4 focuses on a case study, and Section 5 delves into the results. Section 6 concludes the work with a conclusion and recommendations for further research.

2. RELATED WORK

A collection of data, such as supplier, product, and customer information, is referred to as "Master Data" [3]. They are shared throughout most business units, business processes, and IT systems [4]. MDM (Master Data Management) is a broad term that refers to the disciplines, technology, and solutions necessary to produce and maintain consistent and correct master data for all organizational stakeholders [5]. It is a self-contained strategy that allows for maintaining a uniform picture of all master data attributes across all of an organization's data sources, ensuring data quality, consistency, and dependability [6]. MDM should be considered a process improvement program was aiming to discover, assess, and apply processes and technologies to ensure high-quality data for decision-making [7]. MDM, rather than being an application system, is an organizational role.

The dependability of master data is supported by data quality. According to a survey [9] on master data quality management, 80 percent% of businesses believe that poor master data quality substantially influences their financial performance. Companies are forced to comply with an increasing number of rules, necessitating the need for a single source of truth [10]. Organizations' adoption of MDM is heavily influenced by data quality [11].

Data quality may be improved through MDM [12]. On the other hand, many MDM initiatives have been cancelled because all partners in the setup project were not present [2]. As a result, all stakeholders involved in the MDM and data quality process should have common objectives and support one another's efforts [13]. The MDM process can be implemented more effectively if this is completed.

When building an MDM project, the majority of research divides the MDM and data quality link into two topics: (i) the software package, and (ii) the development method.

2.1. MDM as Software Package

Leading software manufacturers have been encouraging the development of MDM solutions in recent years [14]. Although practically all MDM vendors aim to provide a complete solution for MDM, several data quality concerns remain problematic [15]. For example, data profiling is necessary for the early stages of an MDM project to assess the state of master data quality. Despite this, some MDM systems lack these basic functionalities. As a consequence, more partnerships and integrations between MDM and data quality solution suppliers are being developed. Rarely does a software program satisfy all of the criteria [16].

2.2. MDM as a development process

MDM is described in recent research as a combination of technology, business, and data management processes [17]. Several previous MDM models exhibit comparable phases or design areas in starting an MDM project.

The authors in [18] proposed a methodology for developing MDM that included five components: master data systems architecture, master data processes, master data structure, master data quality, and master data governance. When creating an MDM function, the model described in [19] should be taken into account. The seven categories covered were MDM vision, governance, strategy, procedures, technological infrastructure, organization, and measurements. Another approach was provided in [20], which included three categories: organization, process, and MDM systems. These three models focused on identifying the design domains that should be considered when creating an MDM capability. The concept investigated in [21] is aimed at the microfinance industry. It is divided into eight stages and is aimed at tackling data quality concerns. In [15], the author used a process-oriented approach, laying down the steps to be completed correctly. To correctly deploy MDM, the procedure involves eight steps that must be completed. Based on the above work, [22] suggested the following 10 stages for MDM implementation:

• Identifying the need and the goals

• Identifying the company's fundamental data and procedures.

• Data standards must be defined.

• Establishing a framework for governance.

• Metrics for MDM are being established.

• MDM architecture planning.

• Organizing communication and training.

• Defining the procedure for maintenance.

• Creating a development plan for MDM

• Specifying the operational and functional capabilities of MDM software.

The term "data quality" appears in the majority of the studies described above. It does, however, need methods and knowledge management to mature. It is also critical

that all stakeholders in the company have access to the same information to determine the effect data quality may have on the whole business. The earliest stages of MDM development need the establishment of roles and responsibilities for all stakeholders [11].

3. THE PROPOSED FRAMEWORK

We follow the paradigm presented in [22] to create our framework, always keeping data quality as basis in all steps. The resultant framework is divided into numerous phases to accomplish our objective. The Data Quality processes are designed using the Business Process Modelling Notation (BPMN). The adoption of BPMN is justified by the ease with which business analysts and technical developers can comprehend it. As seen in figure 1, the framework is organized into five phases.
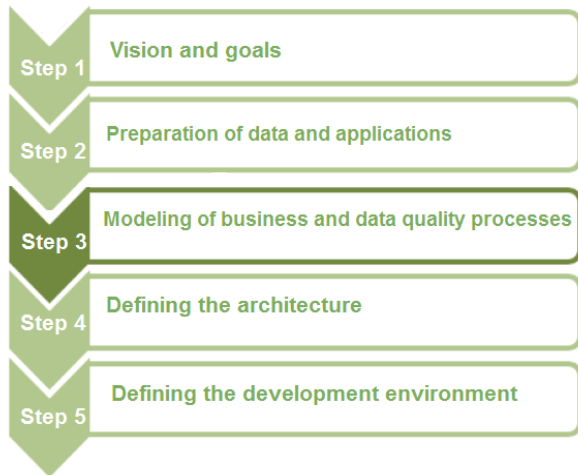

Figure1. The proposed framework structure

3.1 Vision and goals

Assuming that data quality must be a continuous activity inside the business, we will develop the organization's vision and quality methodology concurrently. The following procedures will be put in place:

Communicate with all parties to ascertain the source of data quality issues. This is why business and IT must be included in the quality process by providing an opportunity for them to express openly their assumptions about the conditions in which they anticipate the data to be useable. This stage entails conducting interviews and distributing surveys to those who are directly engaged in the MDM project.

The output of this step will be a table containing the following information: Vision, Goals, and Questions. The table will be constructed using the additional data from step 2. We illustrate this with an example of a business that wishes to maintain positive

relationships with its consumers. Table 1 demonstrates the enterprise's vision.

Table 1. Example of the enterprise vision

Vision	Goals	Questions
Maintain good contact with all employers in the organization	Check that the information to contact the employer is correct (addresses, email, phone, etc.)	**Q1**: is all the employer's information present **Q2:** is the data up to date

3.2 Preparation of data and applications

This stage identifies master data, the applications that generate it, and the applications that consume it. In this instance, it is vital to differentiate master data from other forms of data, such as transactional data. This process requires extensive manual review by data owners and business analysts. This stage results in the creation of a data dictionary to gather master data and its associated metadata.

We must also specify the quality dimensions and associated metrics at this stage. This stage will convert the questions and objectives listed in step 1 into quality dimensions (completeness, timeliness, correctness...), related metrics, and suitable calculation techniques. This table will also have the following columns: Dimension, Measures, Associated calculation technique, and Defined threshold.

After this stage, we will have built the foundation for developing a data quality strategy. Two degrees of detail are defined: the qualities that master data should possess and the data quality metrics that demonstrate how these attributes are assessed.

3.3 Modelling of business processes and data quality processes

Indeed, only managers possess a broader perspective on business processes inside a company. As a result, it is vital to define these processes using diagrams whose representation everybody will share and understand. A data quality process should define how master data quality is collected, calculated, and measured. Additionally, it illustrates the roles and responsibilities of individuals participating in the quality process from IT and business perspectives. The recently released BPMN 2.0 standard enables a considerable degree of detail to be applied to these business processes [23]. Additionally, it is critical that the process modelling be highly adaptable and quickly adaptable to a specific context. This stage

produces a series of diagrams that visually represent the quality process.

3.4 Defining the architecture

MDM is not only a technological function inside the organization. Before addressing technical issues, it is vital to review all preceding processes. Businesses should do a comprehensive study to determine the appropriate strategy. MDM implementations may be broadly classified into three categories of initiatives: Analytical MDM; operational MDM and Enterprise MDM [17]. This stage demonstrates how to use and integrate MDM software inside the organization. The architecture chosen is determined by the company's vision.

3.5 Defining the development environment

Establishing which technologies are required to allow the MDM vision before implementing the suggested framework is vital. The conventional approach, which regards MDM as a technical problem, is abandoned for the duration of the research. Nowadays, businesses must recognize that MDM is a complicated operation that requires sophisticated technological skills. We determine the MDM technologies that best satisfy the organization's software and application requirements in this stage.

4. CASE STUDY

This research aimed to look at data quality issues inside a large enterprise that is trying to implement an MDM project. This company is aware of data quality issues and is attempting to improve its financial capabilities. Consequently, we questioned the company's executives about the data quality problems they had when implementing an MDM project and their thoughts of what causes poor data quality. The answer was that they were primarily organizational. They said that their information system is having technical difficulties. They believe that the quality of data is affected by an employee's expertise. The quality of data is influenced by how users to input information into the system. Erroneous data is permitted to easily flow through business systems, where it goes unnoticed until it is much too late. They also emphasize the complexity of software solutions, which IT specialists can only comprehend. According to one manager, training and documentation are often inadequate since they mainly teach employees how to use the system. He also believes that the most critical factor affecting the quality of master data is a lack of delegation of master data management operations.

For overcoming business issues, a realistic example is given. The findings of the research are divided into three categories:

1. By applying the results of step 1 and establishing a master data dictionary, a common master data definition is shared within the enterprise.

2. The data quality issues mentioned in stages 1 and 2 align with the organization's goals. Managers and IT and business workers will share the same strategic vision if these insights are implemented.

3. All procedures require the early participation of business and IT stakeholders since effective MDM deployment is based on organizational challenges rather than technology. Starting with step 1, where interviews and questionnaires were utilized to determine all stakeholders' intentions, changes are made in step 3 using business process modelling. Multiple BPMN diagrams may be used to define roles and responsibilities further.

5. DISCUSSION

According to the case study, data quality challenges are one of the key motivators for a company to start an MDM project. This is why the data quality procedure is included in most research frameworks and models.

Our results demonstrate that the framework plans the data quality process ahead of time and covers most of the processes, making the process much easier to execute. In addition, our framework lays out the order in which the processes should be completed and their interrelationships. The outputs of one phase become the inputs of the next. However, most research referenced in the literature ignores the interaction between distinct phases since the focal areas are still too broad, as in [15] and [18].

MDM is seen as a joint effort by IT and business departments in our framework. As a consequence, businesses must contribute to the growth of their communities. We argue that skipping this step will lead to misunderstanding. The intersection of organizational and technology design has never been studied before, as shown in [21]. In this case, we propose that the development team consistently offer information that both business and technical management can comprehend. The usage of BPMN bridges the gap between designing and implementing business processes.

We concluded our research that MDM steps should not be data-driven. This indicates that the MDM project is not a software package. Additional MDM solutions that are not included in the software technology must be implemented. As a result, businesses must make many changes that may influence individuals and their work methods and how technology is employed. During the data quality process, our approach promotes integrating individuals who have a clear purpose and responsibility. This group should include business executives as well as IT specialists. He establishes data quality targets and evaluates the effectiveness of the technology solution by

fixing existing data quality concerns and ensuring that the quality level remains high over time.

6. CONCLUSION

This paper addressed the most prevalent data quality challenges in business. A framework has been developed to increase quality throughout the MDM development process. This structure is comprised of five sequential phases. It depicts their interactions with one another. The findings reveal that our system constructs the data quality process in advance, making the procedure considerably more straightforward to implement. Another advantage of our method is BPMN, which bridges the communication gap between business process design and implementation. As shown by the briefcase study, the framework provides essential assistance to enterprises in simplifying and organizing the building MDM programs. Each step of the process of developing a Master Data Management system was regarded as crucial. On the other hand, future research will focus on delivering additional information across the framework's many phases and applying it to a real-world issue.

REFERENCES

[1] R. Volmink-Heikkinen, S. Pecola, Changes in roles, responsibilities and ownership in organizing master data management, in: International Journal of Information Management, 2019, vol 47, pp 76-87, doi.org/10.1016/j.ijinfomgt.2018.12.017.

[2] P. Lienitis, Master data management: its importance and reasons for failed implementations, Doctoral, Sheffield Hallam University, 2020, doi.org/10.7190/shu-thesis-00311.

[3] L. K. W. Fernando and P. S. Hadera, Hybrid framework for master data management, 17th International Conference on Advances in ICT for Emerging Regions (Icterus), 2017, pp. 1-7, Doi: 10.1109/ICTER.2017.8257785.

[4] Y. Liu, H. Liu, F. Yang and X. Chen, Application of Master Data Classification Model in Enterprises, in: 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2020, pp. 1989-1993, Doi: 10.1109/ITNEC48623.2020.9085080.

[5] M. Fleckenstein, L. Fellows, Master Data Management. In: Modern Data Strategy, Springer, Cham, 2018, pp 93-99, doi.org/10.1007/978-3-319-68993-7_10.

[6] E. Baghi, S.Schlosser, Toward a Decision Model for Master Data Application Architecture, in: Proceedings of the 47th Hawaii International Conference on System Science, 2014, pp. 3827-3836, doi: 10.1109/HICSS.2014.475.

[7] R.Vilminko-Heikkinen, P. Brous , S. Pekkola, Paradoxes, Conflicts and Tensions in establishing Master Data Management Function, in: Proceedings 24th European Conference on Information Systems (ECIS), İstanbul, Turkey, 2016.

[8] B. Otto, How to design the master data architecture: Findings from a case study at Bosch, in: International Journal of Information Management, 2012, doi: 10.1016/j.ijinfomgt.2011.11.018.

[9] T. Schäffer, C. Leyh, Master Data Quality in the Era of Digitization - Toward Inter-organizational Master Data Quality in Value Networks: A Problem Identification, In: Enterprise Information Systems Management and Engineering, 2016, Lecture Notes in Business Information Processing, vol 285, Springer, Cham, https://doi.org/10.1007/978-3-319-58801-8_9

[10] B. Otto, Quality and Value of the Data Resource in Large Enterprises, in: Journal Information Systems Management,2015, doi.32.10..1080/10580530.2015.1044344.

[11] F. Haneem, N. Kama, N. Taskin, D. Pauleen, N. A. Abu Akar, Determinants of master data management adoption by local government organizations: An empirical study, in: International Journal of Information Management, 2018, vol 45, pp 25-43, doi.10.1016/j.ijinfomgt.2018.10.007.

[12] D. Loshin, Master data management, in: Morgan Kaufmann publishers, 2010.

[13] A. Haug, J. StentoftArlbjorn, F. Zachariassen and J.Schlichter, Master data quality barriers: an empirical investigation, in: Industrial Management & Data Systems, Vol. 113 No. 2, 2013.

[14] F.Haneem, A.Azmi, Co-dependence relationship between master data management and data quality: A review, in: Journal of Theoretical and Applied Information Technology, 2017.

[15] R.Vilminko-Heikkinen, Data, Technology, and People-Demystifying Master Data Management, Doctoral, Tampere University of technology, 2017.

[16] R.Vilminko-Heikkinen, S.Pekkola, Master data management and its organizational implementation: An ethnographical study within the public sector, in: Journal of Enterprise Information Management, 2017, Vol. 30 Issue: 3, pp.454-475, doi: 10.1108/JEIM-07-2015-0070

[17] M.Allen, D.Cervo, Multi-Domain Master Data Management: Advanced MDM and Data Governance in Practice, in: Morgan Kaufmann publishers, 2015.

[18] A. Cleven, F. Wortmann, Uncovering four strategies to approach master data management, in: Proceedings of the 43rd Hawaii International Conference on System Sciences, 2010.

[19] J. Radcliffe, The seven building blocks of MDM: A framework for success, Gartner research, 2007.

[20] B. Otto, Kai M. Huner, H. Osterle, Toward a functional reference model for master data quality management, in: Information Systems and e-Business Management, 2012, pp 395-425, Doi:10.1007/s10257-011-0178-0

[21] A. Gamero, J. Garcia and C. Raymundo, Reference Model with a Lean Approach of Master Data Management in the Peruvian Microfinance Sector, in: 8th International Conference on Industrial Technology and Management (ICITM), 2019, pp. 56-60, doi: 10.1109/ICITM.2019.8710697.

[22] R.Vilminko-Heikkinen, S.Pekkola, Establishing an Organization's Master Data Management Function: A Step-wise Approach, in: Proceedings of 46th Hawaii International Conference on System Sciences, 2013, pp. 4719-4728, doi: 10.1109/HICSS.2013.205.

[23] OMG, 2016, Business Modeling Specifications (BPMN, BMM, DMN, SBVR, etc.), http://www.omg.org/spec/