Enhancement of an Online Retailer's Data Quality Management Capabilities

Sandro Schweiss

IU International College

Data Quality and Data Wrangling DLBDSDQDW01

Tutor Sahar Qaadan

Abstract

Data quality management plays a crucial role in the success of organizations in today's data-driven world. Ensuring high-quality data is essential for making informed decisions, improving operational efficiency, complying with regulations, and delivering value to customers. This written assignment aims to explore various frameworks and standards for data quality management and their significance in enhancing the data quality capabilities of organizations. These frameworks are then applied to the use case of an online retailer, which struggles with proper data quality management. The goal is to develop an effective plan to assess, improve and implement data quality management in the organization.

Table of Contents

I. Table of Abbreviations

Abbreviation	Meaning
DQM	Data Quality Management
CDO	Chief Data Officer
ISO	International Organization for Standardization
HR	Human Resources
ETL	Extract, Transform, Load

2. Introduction

2.1 Overview

In today's complex business landscape, organizations collect vast amounts of data from numerous sources, including customer interactions, operational processes, and external data providers. However, the data collected is not automatically synonymous with high quality. Data can suffer from issues such as inaccuracies, inconsistencies, incompleteness, and duplication, among others. Such data quality issues can have far-reaching consequences. Poor-quality data can lead to flawed insights, unreliable processes, and detrimental outcomes. Data quality is a critical factor that significantly impacts the success and performance of organizations in today's data-driven world. High-quality data enables accurate analysis, informed decision-making, and effective business operations. Recognizing the importance of data quality, organizations are increasingly focusing on implementing robust data quality management practices to ensure the reliability, integrity, and usability of their data assets. To address these challenges, organizations turn to data quality management frameworks that provide structured approaches for assessing, improving, and maintaining data quality. However, data quality management is not easy to implement, as it requires a thorough understanding of the organization and the ability to develop a data architecture from the ground up. Many businesses, therefore, hire professionals in the field to help them with the management of their data quality. That is why, I was hired as an expert to delve into the data quality management capabilities of an online retail business and help to develop a plan to improve the data quality of the organization.

2.2 Rationale for the Topic

The increasing reliance on data-driven decision-making and the exponential growth of data volumes have brought data quality to the forefront of organizational concerns. Recognizing the significance of data quality in achieving business objectives, organizations are seeking approaches and standards to address this challenge effectively. By examining the different frameworks available for data quality management, one can make informed decisions to enhance their data quality capabilities and mitigate the risks associated with poor data quality, which is an invaluable skill in today’s economy.

2.3 Aim of the Written Assignment

The aim of this written assignment is to provide a comprehensive overview of data quality management, emphasizing the relevance and potential benefits of various frameworks and standards for the online retailer. By exploring these frameworks, to gain insights into different methodologies, practices, and tools, a plan is then developed to assess, improve, and implement data quality management. This assignment aims to develop a plan with the knowledge to assess the suitability of these frameworks for their specific organizational context and develop a robust data quality management strategy.

2.4 Delimitation and Research Methodologies

This written assignment focuses on two primary frameworks for data quality management: the Data Quality Capability Levels and the ISO 8000-61 standard. These frameworks have gained significant recognition in the industry and provide structured approaches for assessing and improving data quality. Additionally, other relevant frameworks and approaches beyond the aforementioned two will also be explored to present a broader perspective on data quality management.

The research methodologies employed for this assignment involve a comprehensive literature review, analyzing academic articles, whitepapers, and official documentation related to data quality management frameworks.

2.5 Organization and Structure

This main body is organized into three main sections. In the introductory section, we provide an overview of the task, its objectives and the overall defintion of data quality management. This is then followed by the next chapter, which delves into the examination of various frameworks and standards for data quality management. The text begins by discussing three primary frameworks: the Data Quality Capability Levels, the ISO 8000-61 standard, and Data Governance. These are the frameworks chosen to develop a data quality management plan. The key components, methodologies, and benefits associated with these frameworks are explored. Then in the third section the text explores the implementation process of data quality management frameworks within the organization. The chapter starts with an assessment of the situation of the online retailer, which is used as a starting point for. Then the plan is proposed for the improvement of the data quality management of the organisation. In the last segment the challenges and methodologies in addressing the plan are discussed. In the concluding section, the key findings and insights from the assignment are summarized. Emphasizing the importance of selecting and adopting suitable data quality management frameworks based on an organization's specific needs and context and comparing the use case to the industry standard.

3. Enhancement of an Online Retailer's Data Quality Management Capabilities

3.1 Overview

3.1.1 Assignment Overview

The following assignment explores a company that tries to enhance the quality of the data it manages on a daily basis. Online retailing constitutes the company's core business, which is supported by a network of on-premises servers hosting several databases. Within this organization, multiple departments exist, and with the exception of the back-office divisions, department heads convene monthly to present the prior month's achievements and forthcoming plans. Despite this reporting structure, interdepartmental communication remains limited, resulting in independent data storage and utilization practices. Various departments employ diverse approaches to store and utilize data related to customers, retail articles, orders, and internal information, including employee records and salary transactions. Data management methods differ among departments, with certain teams storing their data on local computer hard drives while others rely on shared directories hosted by the on-premises server, which simultaneously facilitates the online retailing platform. This decentralized approach has inadvertently created data silos and hindered collaboration across the organization.

To overcome these challenges and enhance data quality management capabilities, the company has sought external expertise. The objective of this assignment is to create a thorough plan that utilizes established frameworks and techniques. By implementing this plan, the company aims to improve data governance, foster interdepartmental communication, centralize data storage, and employ data quality assessment methodologies. The ultimate objective is to harness the full potential of the organization's data assets, facilitate informed decision-making, streamline business processes, and foster innovation (Christian Müller-Kett, 2022).

3.1.2 What is Data Quality Management?

Data Quality Management refers to a set of practices that evaluate, maintain, and improve the quality of data across the organization. It involves processes and tools that ensure that data is accurate, complete, consistent, reliable, and up to date, as well as formatted and used consistently throughout the organization. In today's data-driven business world, data accuracy and reliability have crucial implications. The absence of reliable data can negatively impact business processes, innovation, and decision-making, leading to operational inefficiencies, product failure, high costs, lost sales opportunities, regulatory or financial compliance penalties, and reputational damage. Therefore, enhancing an organization's data quality management capabilities is essential to avoid these issues and improve overall performance (Pipino, Lee, & Wang, 2002).

3.2 Overview of Data Quality Management Techniques and Frameworks

3.2.1 Overview of Data Quality Capability Levels

The Data Quality Management Capability Levels framework provides a structured approach to understanding and improving data quality within an organization. The first level focuses on reliable data processing, ensuring that data meet the demands of end users and incorporating feedback mechanisms. At the second level, a formal and structured approach is adopted to gather requirements and analyze data specifications. This level emphasizes the definition of data processes and objects to deliver the expected output. The third level involves planning data quality and establishing policies and standards to consolidate existing processes. It serves as a control mechanism for levels one and two. The fourth level assesses the overall effectiveness and efficiency of data quality management activities, identifying areas for improvement and ensuring alignment with standards. Finally, the high-end capability level five aims to improve data quality through root cause analyses and raising awareness throughout the organization, fostering a more data-driven culture. This framework enables organizations to identify areas of improvement and take progressive steps toward achieving higher data quality (Loshin, 2011).

3.2.2 The ISO 8000-61 Standard

ISO 8000-61 is a comprehensive standard for data quality management that provides practical guidelines for improving the data capability level of an organization. The standard encompasses four core processes: data quality control, planning, assurance, and improvement. These processes are supported by two high-level processes: data-related support and resource provision. Data quality control ensures that data requirements are met, especially in scenarios where data or systems are frequently updated. Data processing focuses on creating data objects that support decision-making. The standard emphasizes identifying areas where data processing takes place and understanding their potential impact on data quality. The provision of data specifications involves developing artifacts and data specification standards in collaboration with domain experts. The focus is on the syntax, semantics, and pragmatics of the data, tailored to the needs of end users. Data quality control defines methods, tools, and techniques to test and validate data against predefined specifications. This process aims to standardize methods and tools across the organization and includes manual auditing efforts. Data quality planning involves formally defining and agreeing upon data quality requirements and objectives. Data quality assurance includes four processes: review of data quality issues, provision of measurement criteria, measurement of data quality, and evaluation of measurement results. These processes aim to identify data quality insufficiencies, establish measurement criteria, monitor progress over time, and provide recommendations for improvement. Data quality improvement comprises three processes: root cause analysis and solution development, data cleansing, and process improvement for data nonconformity prevention. These processes analyze

root causes, develop solutions, reduce false data entries, and optimize processes. Lastly, data-related support processes, such as data architecture management and data transfer management, ensure a suitable data infrastructure is in place for effective data quality management (“ISO 8000-61,” 2023)

3.2.3 Data Governance to Enforce Policies and Procedures

The role of data governance is of utmost importance in ensuring the adherence to policies and procedures concerning the management of data. While data quality management focuses on improving the quality of data, data governance takes a broader perspective by establishing formal policies and standards across an organization. It encompasses strategic activities such as defining policies and ensuring their enforcement through centralized governance while delegating execution. Data governance and data quality management overlap in areas such as the creation of policies and standards. Both areas recognize the importance of data stewardship, which involves the role of a data steward responsible for overseeing data activities and delivering fit-for-purpose data to users. Data stewards, as specialists in data governance and data quality processes, develop process frameworks, policies, and guidelines while ensuring alignment with the main principles of data quality management. In addition to data stewards, other roles in data management include the Chief Data Officer (CDO) who focuses on strategic aspects of data utilization and identifying business opportunities, and data owners who are responsible for specific datasets. The ownership of datasets may involve multiple stakeholders due to the combination of various processes and inputs. (Samitsch, 2015).

3.3 Developing a Plan for Enhancing Data Quality Management

3.3.1 Assessment of the Current Data Quality Capability Level

The first step to developing a plan is to conduct a thorough analysis of the current data landscape within the organization. Upon assessing the existing databases, data storage practices, and data usage across departments, it becomes apparent that there are significant challenges in terms of data quality and management. The company operates with multiple databases deployed on on-premises servers to handle online retailing. While back-office departments seem to have some level of data organization, other departments operate in silos with limited communication and disparate methods of storing and utilizing data. Data related to customers, products, orders, and internal processes are scattered across individual computers and shared directories. This fragmented approach has hindered business processes and stifled innovation due to the lack of insights gained from the accumulated data. Based on the previous elaboration of data quality management techniques, it can be concluded that the organization's data quality capability level is at level one. Due to lacking any kind of standardized approach, no centralization of data, minimal quality controls, and ad hoc architecture make it practically impossible to gain any kind of insight into the data. Therefore, the organization's data quality management must be built from the ground up.

3.3.2 Development of a Data Quality Management Plan

To complete the level one data quality capability level, the focus is put on exactly one key aspect, namely, providing the end users with data that satisfy their needs in the decision-making process. This means that in this step the focus lies in identifying the areas where data processing takes place and how it influences data quality. The objective is to gain a comprehensive understanding of the data processing flow within the organization, including data inputs, transformations, and outputs (King & Schwarzenbach, 2020). In the case of the online retailer, the areas of interest are customer data, which includes personal information, purchase history, and demographic details. Product data, covering descriptions, specifications, pricing, and inventory levels. Order data is vital for order fulfillment, inventory management, and sales. And at last internal Data, which is related to employees, salary transactions, and other operational aspects. Then it is critical to establish what the data flow in the organization should be, in other words from where the data is pulled from and to what departments the data should be supplied to. Since this is an online retailer, the customer makes an order on the official website of the store. The order is then processed by the online retailing platform of the organization and stored on the on-premises server of the organization. The customer data, order data, and to an extent product data, are included in the order when buying on the website. Customer data should then be supplied to the marketing department, for different marketing campaigns and to identify target audiences. The sales department needs the order and product data, which is also provided by the supplier, to manage inventory and fulfill customer orders efficiently. The other kinds of data like internal data are however stored on hard drives of local computers and some on shared directories making the data incomplete and unusable. Internal Data like employee information and salary transactions need to be supplied to the HR department. The overall data about finances need then to be supplied to the finance department. All of this information should then be supplied to the executive of the organization in an even more distilled version to perform higher-level decision-making. The outcome of this stage is the creation of centralized storage and the definition of informal data quality objectives. The organization should convert its on-premise server to a centralized data repository that serves as the foundation for data integration. This repository would consolidate data from various sources and departments. By centralizing data storage, the organization can eliminate data silos and minimize duplication, which, allows for improved data accessibility, visibility, and consistency.

Achieving level two in data quality capability involves the formal and systematic analysis of data specification. Once the centralized data repository has been established it facilitates a more specific understanding of the exact type of data that is required. During this step, specific work instructions are created for the data process. This includes a close collaboration with the domain experts like HR, sales, and marketing departments to define the exact attributes the data should have and how the data is displayed. This is then communicated with the operational departments of the business,

who process items like orders, supplies, or products, which in turn produce the data. The outcome of this stage is a formal test for deviations. Here the planned data standardization is tested against the actual outcome of the data (King & Schwarzenbach, 2020). The goal is to create standardized methods and tools across the organization to maintain a certain level of quality as automatically as possible. The result is a formal description of the overall process and the specifics of the data.

The attainment of data quality capability level three requires a comprehensive plan for data quality that encompasses various segments of the organization. Building upon the foundation established in previous levels, this stage focuses on planning and consolidating existing processes through the development of policies and standards to ensure effective control mechanisms. From an organization-oriented perspective, several key aspects need to be addressed. It begins with the development of specific requirements from all stakeholders. This step helps align data quality efforts with the needs of the organization and its stakeholders. Based on the requirements management, a data quality strategy is developed to outline a roadmap for increasing data quality capabilities. The overall objective of the business can change, depending on business goals. A common goal in the online retail market is increasing the number of conversions. Another important aspect is the management of data quality policies. Guidelines and procedures are created to provide clear instructions on how to act in various scenarios, ensuring consistency in data handling practices. The big picture is analyzed, and smaller processes are improved. An outline is created to guide the delivery of data quality standards and policies to the entire organization. The third level also takes the first step into data governance. This is done by either transforming existing roles or creating new roles that are specifically dedicated to data quality management. This is done by acquiring personnel with the necessary knowledge and skills to execute data quality management in line with the specified policies, standards, and procedures. The SFIAplus framework can be employed to assess the adequate skill level required for each role, ensuring that the right individuals are placed in positions that align with their capabilities. These roles focus on driving data quality initiatives and ensuring their effective implementation within the organization. This should result in identifying suitable data owners, which are responsible for specific databases, and facilitating the deployment of guidelines and procedures. A major focus is also dedicated to the technical perspective, in the form of data-related support. Data architecture management involves creating data models to describe the data in different stages of the lifecycle, enabling a better understanding of the overall data architecture. This step provides a more sophisticated approach to how various different devices are linked to the system. From these devices ETL pipelines are then created, to facilitate data accessibility and integration across systems. These pipelines streamline the movement of data, ensuring its availability when and where needed. Once the ETL processes have been implemented, data operations management comes into play to implement infrastructure maintenance procedures, such as database backups, system optimization, user access management, and database management. Data security management is

also critical at this stage, involving the implementation of security standards and norms to safeguard data assets. Adhering to recognized frameworks such as ISO/IEC 27000 ensures that appropriate security measures are in place to protect sensitive data. These practices help ensure the smooth operation and performance of data-related systems (King & Schwarzenbach, 2020). The achievement of data quality level three is the most significant step in the organization, due to being the step, which transforms the business on an organizational, technical, and human level. The end results of this stage are defined sets of standards and procedures, as well as blueprints for the ETL process for all participants. Employees should not only have a holistic understanding of the process, but an exact understanding of which actions trigger progress in the ETL pipeline. In addition, there should be a progressive mindset shift towards data quality understanding in the organization, with the first introduction of data governance mechanisms in the form of data owners. In the case of the online retailer, data stewards should also be deployed. They work closely with different departments to address data quality issues and would improve the communication problem.

Level four of the data quality capability plan focuses on enhancing the overall effectiveness and efficiency of data quality management activities within the organization. This stage involves several key components that contribute to the continuous improvement of data quality processes. To begin, a comprehensive review of data quality issues is conducted to identify general data quality insufficiencies at the outset of the improvement process. This step helps uncover specific data quality issues that can be prioritized and addressed during the subsequent data quality improvement phase. In order to effectively measure and track data quality progress, it is crucial to establish appropriate measurement criteria. The measurement criteria enable the organization to evaluate data quality over time and assess the impact of improvements made. By measuring data quality against these established criteria, trends and patterns can be identified, providing valuable insights into the root causes of data-related issues. The next aspect is the measurement of both data quality and process performance. A framework is developed to measure the performance of data quality as well as the underlying data processes themselves. This evaluation helps identify areas that require further attention and improvement. The evaluation of measurement results is a critical step at this level. It involves analyzing the implications derived from the measurement of data quality and process performance (King & Schwarzenbach, 2020). The outcome of this step is a set of measurement criteria against which the identified data quality issues are measured against. This comprehensive assessment results in a report that showcases the progression of data quality and highlights the performance of each data object alongside its associated processes. Based on the findings, recommendations are provided to enhance several data quality processes. This process culminates in the creation of action items, such as a data quality management health checklist, which outlines specific steps to be taken to address the identified data quality issues. At this stage, a CDO may be deployed to oversee the organization's data quality management initiatives.

Level five, the last level of the data quality capability plan represents the high-end capability level, focusing on improving data quality through comprehensive root cause analyses. This stage is centered around identifying the sources that create errors within the data. The first key component is the root cause analysis and solution development. This involves conducting a thorough analysis of the root causes behind data errors. It takes into account people, processes, technologies, and organizational structures to identify the underlying issues rather than treating symptoms. Techniques such as Ishikawa (or fishbone) diagram and the 5-Whys can be utilized to dig deep into the root causes of data quality issues Data cleansing is another crucial aspect of level five. The focus is on reducing the number of false data entries to prevent the propagation of errors throughout the data stream. Data cleansing involves thorough validation and verification processes to ensure that the dataset meets the requirements for accuracy, consistency, and coherence. Process improvement for data nonconformity prevention is a key consideration here. This step evaluates all defined processes and aims to identify weaknesses and opportunities for improvement in preparation for the next iteration. The focus is on optimizing processes at each capability level with a specific emphasis on identifying underperforming processes. Through process improvement initiatives, organizations can proactively address data nonconformities and enhance the overall effectiveness and efficiency of their data quality management practices (King & Schwarzenbach, 2020). The outcome of level five is the development of effective solutions that address the identified root causes and mitigate the occurrence of errors. Then by implementing data cleansing techniques, the online retailer can enhance the overall quality and reliability of their data. As a result, the organization will experience enhanced data quality, improved decision-making processes, and increased overall efficiency in utilizing data as a strategic asset.

3.3.3 Implementing the Change in the Organisation

Implementing change in data quality management within the online retailer requires an understanding that it is an iterative process, not achievable in a single project but rather through multiple projects. Transitioning from one data quality capability level to another requires a systematic approach and careful planning. It is essential to establish a clear roadmap that outlines the steps and milestones for each transition. This involves identifying the necessary resources, defining the timeline, and aligning stakeholders' expectations. It is important to consider that this is not only a change in the organizational structure but also the mindset of all stakeholders. Additionally, it is important to establish a change management framework that includes effective communication, training, and support mechanisms to facilitate a smooth transition. Continuous improvement practices should be embraced, fostering a culture of learning, feedback, and adaptation to ensure sustained progress, culminating in a department for data quality management. By following these best practices, the online retailer can navigate the transition process effectively and drive continuous improvement in their data quality capabilities (King & Schwarzenbach, 2020).

4. Conclusion

Data quality management is crucial in today's data-driven business world, as it ensures that the right data that are fit for the purpose is delivered to the intended user. If this is not the case, the consequences can be so severe that an organization could never evolve beyond a certain stage (Wang & Strong, 1996). The assignment task highlights the exact challenges faced by an online retailer in managing the quality of its data and explores the steps necessary to enhance data quality management capabilities. The organization operates with decentralized data storage practices, resulting in data silos and limited collaboration between departments. To resolve this kind of problem, there are many frameworks that can be deployed like the “Total Data Quality Management” framework, “Data Governance” framework, “DAMA Data Quality Dimensions” and many more. For this assignment, the “Data Quality Management Capability Levels” framework was determined to be the best solution. It provides a structured approach to understanding and improving the previously discussed problem. It encompasses different levels, starting from the essentials like reliable data processing, and progressing to the most granular level of data processing like root cause analyses and a data-driven culture. The ISO 8000-61 standard offers practical guidelines for data quality management, covering processes such as data quality control, planning, assurance, and improvement. Data governance also plays a crucial role in enforcing policies and procedures related to data management, ensuring consistency and alignment across the organization. The developed plan for enhancing data quality management first involves the need to assess the organization's current data quality capability level and from there employ strategies to work upward toward higher data quality capability levels. In this case, the organization operates at level one, with fragmented data storage and limited quality controls. The plan, therefore, focuses on working from the lowest level toward the highest. Starting with the focus on the basics like centralizing data storage, to final steps like establishing its own department for data quality management and a data-driven organizational mindset. Comparing the findings of this case study with other studies and best practice examples from the literature, it is evident that data quality management is a critical aspect of organizations operating in a data-driven environment. The challenges faced by the online retailer in this assignment are not unique, as many organizations struggle with decentralized data storage and limited collaboration. Best practices emphasize the importance of centralization, standardization, and the involvement of data stewards and data owners to drive data quality initiatives. Another aspect that holds many organizations back is the wrong mindset. Many businesses do not operate with a data-driven mindset, which can severely affect the decision-making process and may keep the organization from evolving (Bennett, 2023). Data quality management has become an invaluable component of the modern management process and is essential to stay competitive in a data-driven world.

V. Bibliography

Bennett, S. (2023, August 6). Data Quality Statistics. *WebinarCare*. Retrieved from https://webinarcare.com/best-data-quality-software/data-quality-statistics/

Christian Müller-Kett (2022). Task: Written Assignment.

ISO 8000-61: 2016(en), Data quality - Part 61: Data quality management: Process reference model (2023, June 23). Retrieved from https://www.iso.org/obp/ui/en/#iso:std:iso:8000:-61:ed-1:v1:en

King, T., & Schwarzenbach, J. (2020). *Managing data quality: A practical guide*. Swindon, England: BCS, The Chartered Institute for IT.

Loshin, D. (2011). *The practitioner's guide to data quality improvement*. Burlington, MA: Morgan Kaufmann.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211–218. https://doi.org/10.1145/505248.506010

Samitsch, C. (2015). *Data Quality and its Impacts on Decision-Making*. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-08200-0

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099