End-to-End Data Quality: Insights from Two Case Studies

M. Redwan Hasan and Christine Legner
University of Lausanne, Switzerland

Abstract: Maintaining high data quality in organizations have become indispensable. In the past, companies largely concentrated their data quality efforts on a single point in the information supply chain – focusing either on master data quality or on information products. As they start repurposing data and leveraging it for more advanced and complex use-cases, they need to proactively manage data quality in an end-to-end approach. Leveraging insights from two case studies, this paper analyses two different, yet complementary approaches to end-to-end data quality management, namely first-time-right approach and use-case driven approach. The findings highlight that end-to-end data quality management relies on common principles but can start from either side of the information supply chain – either through a use-case or data entry point at the source.

1 INTRODUCTION

The amount of data has been increasing at an exponential rate. A survey of more than 2000 business and IT managers found that data is expected grow nearly five times by 2025 with 57% fearing their businesses would not be able to keep up with such massive volumes (BusinessWire, 2020). Data flows through an information supply chain which processes and transforms it into an information product for the use of data consumers (Wang, 1998). When more users and systems interact with the data in the process, this leads to a higher possibility of dilution in the quality of data (Taleb, Serhani, & Dssouli, 2018). Thus, data quality needs to be embedded throughout this journey of becoming an informational output, rather than just focusing on the quality at a single point in the chain. This calls for an end-to-end view, that connects the different users, systems and processes interacting with data in the information supply chain – facilitating a proactive and ongoing exchange of details on identification and correction of poor data quality when it manifests. Such view allows for a better awareness and stronger control which is vital for data quality (Jones-Farmer, Ezell, & Hazen, 2014). It closes the loop in two ways: First, by connecting relevant entities that constantly communicate and proactively ensure data quality (Krishnan, Haas, Franklin, & Wu, 2016). Second, by establishing continuous improvement cycles, as suggested by data quality management methods, such

as the seminal Total Data Quality Management (TDQM) approach (Wang, 1998) and the Define, Measure, Analyse, Improve, Control (DMAIC) cycle from Six Sigma (de Mast & Lokkerbol, 2012).

Although few recent papers (Byabazaire, O'Hare, & Delaney, 2020; Taleb et al., 2018) stressed the need for end-to-end data quality throughout the data pipeline, they are mainly centred around big data. Existing data quality research, on the other hand, has mainly looked into barriers for master data quality (Haug & Arlbjørn, 2011; Loshin, 2010), measuring master data quality using a cockpit (Otto, Ebner, & Hüner, 2010), improving data quality using master data management (Hikmawati, Santosa, & Hidayah, 2021) and controlling data quality at source (Singh & Singh, 2010). Other authors studied information product or data product quality (Machado, Costa, & Santos, 2021; Parssian, Sarkar, & Jacob, 2004), but were limited to only relational databases using certain data quality dimensions. We conclude that extant literature has considered data quality at different points in the information supply chain, but that we lack empirical studies to better understand end-to-end data quality management within the realities of data flows in enterprises. To address these gaps, we propose the following research question:

How do firms implement end-to-end approaches to manage the quality of their data?

To analyse data quality management in a real-life context, we opted for multiple case studies (Yin,

2003). We selected two multinational companies that we consider critical cases (Paré, 2004). Both have implemented data quality from an end-to-end perspective but pursue different approaches. Both companies have matured data management teams with strong data governance and put specific emphasis on establishing closed loops that enable continuous data quality improvements. In this paper, we analyse both approaches with their commonalities and differences through the lens of DMAIC cycle. The latter outlines the main phases for managing quality improvements in organizations (Montgomery & Woodall, 2008).

From our within- and cross-case analysis, we find that both companies address all phases of the DMAIC cycle, but their approaches highlight different initiation points, root-causes & improvement methods. While one company implements master data quality at the source to support an increasing number of business processes at global scale, the other leverages a use-case driven approach that improves data quality for a small set of relevant data attributes for high-priority analytics and operational use-cases. By identifying patterns towards end-to-end data quality, our findings contribute to existing data quality literature (Otto & Österle, 2015; Zhu, Madnick, Lee, & Wang, 2014) and provide a starting point for future research regarding how data quality can be ensured at each stage of the information supply chain especially when organizations are increasingly collecting and utilizing different forms of big data.

In the next section, we review the data quality literature. Then, we define the research gap and discuss the research methodology. Next, we introduce the case studies and perform the within- and cross-case analysis. Finally, we present our conclusions, limitations and outlook on future research.

2 BACKGROUND

2.1 Defining Data Quality

Data quality is most often defined in terms of data's "fitness for use" (Tayi & Ballou, 1998). Thus, data quality is likely to vary among people and functions based on the tasks they seek to address. For instance, Wang & Strong, (1996) produced an extensive initial list of 179 data quality dimensions, 15 of which were identified for practical use and were categorized into four data quality hierarchies – intrinsic, contextual, representational and accessibility. Various data quality dimensions, such as accuracy, volume, completeness, timeliness and trustworthiness are

highlighted in various papers (Klein & Lehner, 2009; Metzger, Chi, Engel, & Marconi, 2012). To assess data sources, accuracy, validity and credibility were underscored as required dimensions (Barnaghi & Sheth, 2016). However, data quality dimensions required to measure data quality varies for different data types (Batini & Scannapieco, 2016) and also for various application domains and data sources (Batini, Rula, Scannapieco, & Viscusi, 2015). While existing studies mostly elaborate on different data types and the relevant data quality dimensions, they do not consider the data flows in enterprises and where data quality should be measured.

2.2 Data Quality Management as Continuous Improvement

Another stream of the data quality literature focuses on developing and applying various frameworks to manage the quality of data (Batini, Cappiello, Francalanci, & Maurino, 2009). For instance, the seminal work by Wang (1998) on the TDQM framework encourages a product perspective on data and provides four stages to ensure end-to-end quality improvement process. The Total Information Quality Management (English, 1999) approach focuses on the management implications of consolidating operational data into data warehouses. To evaluate web-based information using tools, the Information Quality Measurement approach (Eppler & Muenzenmayer, 2002) outlines assessment planning, configuration, measurement and follow-up activities as steps. The Activity-based Measuring and Evaluating of Product Information Quality (Su & Jin, 2007) assesses data quality in manufacturing companies that produce physical products. Most of the frameworks above are designed to meet data quality in a specific context and are not general-purpose in nature. Seminal work like TDQM which is argued to be general-purpose (Batini et al., 2009) lacks the control step which is crucial in ensuring high data quality (Jones-Farmer et al., 2014).

The DMAIC cycle from Six Sigma is widely used for process and quality improvement. It provides a structured and general problem-solving guideline (Montgomery & Woodall, 2008), allowing organizations to better understand the complexities behind initiatives such as data quality. The DMAIC cycle comprises five phases (Smętkowska & Mrugalska, 2018):

a) *Define* – The define phase starts with the identification of the data quality problem, its business impact and resource needs.

b) *Measure* – The measurement phase defines the metrics that are scored in order to quantify the existing data quality issues.

c) *Analyse* – The analysis phase interprets the metrics results and identifies the root causes to the data quality problem.

d) *Improve* – The improvement phase puts actions, techniques or solutions in place to fix the data values or change processes.

e) *Control* – The control phase checks whether the improvements are sufficient and monitors deviations from the objectives.

The primary principle of DMAIC is to establish a continuous cycle of identification and improvement of data quality-related challenges that feeds into the next iteration. By doing so, it closes the loop. As the phases take place sequentially, it leads to a continuous evaluation of the data quality initiatives within the loop – leading to a sustainable perpetuation of the data quality tasks (Montgomery & Woodall, 2008).

2.3 Research Gap: End-to-End View

Despite the ongoing debate on data quality, we observe a void of literature that captures data quality from an end-to-end perspective. The need for end-to-end view is exacerbated by emerging analytical use-cases that are increasingly playing a key role in creating business value. Such use-cases require a lot of data from multiple sources within the organization. These data have been collected, stored and transformed in numerus ways by various teams. Hence, while running the analytics use-cases, data consumers could lack the insight whether the right data with appropriate quality is being used – implying a lack of overview into the journey of data in the information supply chain. Hence, an end-to-end view will provide clarity regarding when, where and how data quality was hampered and how to effectively fix and sustain it. To address these gaps, we call for empirical studies investigating how end-to-end data quality has been put in practice – enriching our insights about the different and similar ways in which organizations conduct the end-to-end implementation with the singular objective of improving data quality. Owing to the huge surge of data and advanced analytics use-cases, this practical understanding is crucial to manage the ever-evolving data quality requirements and challenges because organizations are becoming more data-driven. Therefore, this study will also lay groundwork to guide organizations to adapt and scale their data quality initiatives based on changing data needs in their respective business environments.

3 METHODOLOGY

To address our research question, we opted for a case study research design (Yin, 2003). Case studies provide the opportunity to study the phenomenon of interest in a naturalistic setting and understand it within real-world context (Benbasat, Goldstein, & Mead, 1987). Evidences garnered from several case studies are often more compelling, regarded as more robust and helps derive analytical generalizations (Yin, 2003). We opted for two case studies, because this significantly improves the analytical benefit and the conclusions arising would be much stronger than compared to a single case study. Hence, "… having at least two cases should be your goal" (Yin, 2003, p. 54). We selected two companies as critical cases (Paré, 2004), that have implemented end-to-end data quality but use different strategies to attain this goal. We used the following criteria to guide the selection of the companies: First, the two companies are major players within their respective industries and often feature in the Fortune 500 list. They have significant global presence and operate across multiple continents. Secondly, both are large organizations with strong experience in data management practices and emphasis on end-to-end implementation. Thirdly, the two companies received significant recognition due to their innovative data quality management approaches. They had been shortlisted as finalists for good data quality practice award, after being assessed by jury of international data management experts comprising of academics and practitioners. Therefore, being data-driven allows them to leverage existing data and processes in order to create key insights which allows them to efficiently run operations globally. Due to their global presence, it is particularly challenging to improve the quality of data in an end-to-end manner – providing a setting to empirically study different data quality management approaches with the same goal. The overview of the case companies is given in Table 1.

We collected data through the following primary and secondary sources: The application documents of the two companies submitted for the award were initially analyzed. From this analysis we got a first

Table 1: Overview of the case companies.

Company and (Industry)	Revenue/Number of employees	Data Quality Improvement Approach	Goals of the Data Quality Improvement Approach	Achievements
FashionCo / (Fashion and Sportsware)	$1-50B/60,000	Use-Case Driven Data Quality Management	Improve the efficient application of vital use-cases (e-commerce, sustainability, etc.) by fixing the quality of data in a reduced set of relevant data attributes	Improvement of 5 use-cases with 3 feedback loops. Up to 40000 data defects were fixed in these use-cases.
ChemicalO (Specialty Chemical)	$1-50B/40,000	First-time-right data lifecycle process	Supply high-speed and first-time correct business partner data through harmonization of various data lifecycle processes	Process lead time improved by 66%. First-time-right rate of the data reached 80% from 40%.

overview of their approaches. The companies also included video demos to show certain aspects of their data quality approach and provide details. Moreover, we participated in their final presentations during the award ceremony and in the subsequent discussion of the cases with data management experts.

As part of the within-case analysis, we mapped an initial breakdown of the case data against the DMAIC cycle. This allows "the unique patterns of each case to emerge" (Eisenhardt, 1989, p. 540) and eventually helps lay foundation to gain deeper insights and rich familiarity in the selected cases. Subsequently, to grasp the patters across the cases, we perform a cross-case analysis "to go beyond initial impressions, especially through the use of structured and diverse lenses on the data" (Eisenhardt, 1989, p. 541). This better elucidates the commonalities and differences between the two end-to-end data quality approaches, enhancing the reliability and accuracy of the analysis. Also, new insights could be found that might not have been possible through a simple within-case analysis because cross-case analysis deepens the explanation and understanding of the identified patterns (Miles & Huberman, 1994).

4 CASE OVERVIEW

4.1 FashionC

As global fashion and retail company, FashionC faces the challenge of a fast-changing seasonal product portfolio with around 100,000 active products and several 10,000 new products per season. FashionC traditionally sold via retail channels, but e-commerce and direct-to-consumer channels are playing an increasingly important role leading to an increase in the amount of data FashionC is producing.

The requests for resolving data quality issues for both analytical and operational use cases were high.

Based on learnings from 13 high-priority use-cases, FashionC developed a Use-Case Driven Data Quality Management approach, which aimed at establishing sustainable links between data consumers and data creators. The key elements of the approach can be summarized as follows: The trigger is a data quality improvement request from the data consumers for business-critical use-cases. The data quality team identifies relevant data attributes with quality issues. Depending on the use case, these are typically very few attributes (up to 4), for which a definition and shared rules need to be defined in a first step. Only with these definition and rules, data quality can be measured and the issues can be made transparent to all stakeholders. This allows data change requests to flow faster and be implemented within a short time due to a direct connection of data consumers and data producers. The data quality requests are sent to the data producers through existing platforms such as MS Teams and MicroStrategy data quality dashboards. The data producers correct the data issues and provide confirmation back to the data consumers.

For instance, a data consumer identifies problems in 'sustainability and ethics compliance validation'. They look at the business rules that are behind the use-case such as 'material data should be compliant to SEC' and 'SEC, product hangtag, F&B must comply'. These rules then define the relevant data attributes needed to run the use-case, such as 'hangtag', 'features and benefits', 'technology concept' or 'material composition'. These data attributes fall under the jurisdiction of the material team and they are informed to change the data values – creating a closed loop by connecting directly with the responsible person at source and fixing data quality issues quickly.

Within 3 months, FashionC aims to solve the data quality issues existing in a selected use-case. Until now, FashionC's data quality team worked on 5 use-cases and established 3 such feedback loops. For one specific use case, this resulted into fixing defects

worth 1 million Euros and 80% less effort in correction and escalations. The overall success is communicated via internal channels to garner interest around data quality. Furthermore, it builds credibility of the data quality team and encourages further identification of use-cases to fix.

4.2 ChemicalO

Being a global specialty chemicals company, ChemicalO's adaptation of a new corporate vision of 'profitable growth' through innovative processes have put the usage of data analytics and new technologies into focus. The motivation driving this vision is high innovative capability, consistent end-to-end view of data and cost efficiency & reliability.

The data quality issues manifested due to the decentralized data entry into several isolated systems. This led to the creation of business partner data that did not pass any formal quality checks but is manually entered. Such isolated entries led to inconsistencies and quality checks on the data were required before it could be used in important processes. To do so, an assessment of the criticality of the data is performed. If the data was highly critical, a manual workflow was separately installed to fix the data. On the contrary, less critical data were fixed by the data consumer on their end. These segregated steps made the whole process slow and inefficient, ending up reducing the quality of the data.

To address this issue, ChemicalO introduced the CuVenSa – ChemicalO's journey to a touchless first-time-right data life cycle process. More precisely, the company developed a user-friendly data self-service to consolidate all the sub-processes, namely creation, extension, change and deletion, that were needed to manage the business partner data. The prior manual checks were automatized by connecting with external data sources and internal databases. This allowed ChemicalO to utilize trustworthy external data to run checks on the business partner data quality internally and fix them through few clicks rather than waiting for the internal process to learn about it first.

For instance, for payment fraud detection, prior confirmation of vendor bank information was done through manually communicating with a reliable contact at the vendor company. With CuVenSa, ChemicalO could approve bank information thanks to external data from an inter-enterprise shared data pool. Simultaneously, ChemicalO built up an internal database with bank details in order to further complement this pool. Together, ChemicalO achieved 60% automatic bank validation from 10% without contacting the vendor at all.

Following a time span of 12 months, ChemicalO's data quality team was able to deploy operational tasks 66% faster. In addition, over 500 Person Days per year were saved for over 3000 data requesters and maintainers. Better data also led to saving of efforts worth 1.7 million Euros and 17000 hours for tax data audits. The overall success was communicated using short videos through organizational channels in order to build credibility of the data quality team. More importantly, it led to building trust in the usage of external data to improve the quality of data.

5 CROSS-CASE ANALYSIS

5.1 Define Phase

In the case studies, companies have taken a contrasting position in order to define the data quality problems. More specifically, FashionC decided to focus on data quality problems for high-priority operational and analytical use-cases such as e-commerce and sustainability. Therefore, they adopted the strategy to drive data quality problem identification from the side of data consumers. These use-cases are driven by new data requirements, often a combination of existing sources, and concern specific data attributes. On the contrary, ChemicalO's approach embraced the view that data quality issues must be identified and fixed at source and concentrated on one of their most fundamental master data entities, i.e. business partners data. Business partner data flows into downstream systems used by many data consumers for master data or other processes. If the data quality issues linger at source, it will cascade down into global business processes and create performance issues. This highlights that the end-to-end approaches to improve data quality can start from either side of the information supply chain.

5.2 Measure Phase

In this phase, the data quality issue is quantified in order to better understand the nature of the problem. ChemicalO's data quality problems appear at the source system for the entire dataset pertaining to business partner data. As the dataset would be used in global processes, the strategy was to ensure data quality for all the data instances present in all the data attributes. Since business partner is an established master data object, existing data validation rules were applied to measure whether the data values exist and comply with the rules. These rules helped check whether the data were inaccurate, missing or

Table 2: Two approaches to end-to-end data quality.

	FashionC: Use-case driven data quality	ChemicalO: First-time-right data quality
Trigger	Increasing number of data use-cases in operations and analytics area need combination of different data sources	Inconsistencies in business partner data that impact global business processes
DMAIC Cycle		
Define	New data quality requirements in (few) relevant attributes for business-critical use-cases	Data quality problems for established master data objects ("single source of truth")
Measure	The data quality is measured in the specific datasets based on shared definition and business rules	The data quality is measured in the source system for the entire dataset based on data validation rules
Analyse	The root cause of poor data quality is either wrong data capture, logic issues in the source systems, system integration or wrong data usage	The root cause of poor data quality is primarily due to defects in data creation
Improve	Automated feedback loops between data consumer and data producer to create transparency about the data quality problem and allocate responsibilities to fix them	Harmonize and automate the data life-cycle processes through self-service application and using external data connections
Control	Stakeholder-oriented data quality dashboards to constantly provide overview of the data relevant to them	Process-related KPIs dashboard to create transparency across the whole organization

inconsistent. For instance, P.O. Box and location data of suppliers must be consistent as it has vital tax implication for a global company like ChemicalO. On the other hand, FashionC had to start measuring quality for datasets that combined different data sources. They concentrated only on the data attributes that are relevant for specific use-cases. The problematic use-cases are re-engineered back towards the data source and the data quality team had to work with data consumers and producers to create a shared definition of the business rules and attributes that power it. Once identified, the business rules help measure the data quality in only those attributes that matter and the data values that do not meet the rules are checked. For instance, for 'customer inactivity monitoring', three relevant business rules were identified and inactive customers were defined using financial transaction data attributes. This allowed to highlight around 20000 customers that were inactive.

5.3 Analyse Phase

The analysis phase concerns the understanding of the root cause of the data quality problem. ChemicalO had initially implemented multiple data lifecycle processes in a way to respond to data consumers' need of accurate and high-speed data. When it came to entering business partner data into the system, only certain data attributes that were deemed to be critical went through extra steps of manual quality checks whereas other non-vital data were readily entered without much control. For the latter, the onus was on the user to identify data problems and fix them on the go. This fragmented approach made it difficult to

manage quality of incoming data at source in a harmonized manner. Hence, the root cause of poor data quality was identified at the data creation point where data was wrongly entered through different processes. On the other hand, FashionC's root cause for poor data quality was difficult to locate. As a use-case typically requires multiple data items, analysing data quality requires a re-engineering effort by going back in the information supply chain to see what happened to the relevant data attributes. The root cause appears not only during data capture by data producers but also due to system integrations and usage challenges in the supply chain. As a result, data quality gets diluted in the journey.

5.4 Improve Phase

In order to deal with the root cause and improve the data quality problems, ChemicalO focused on data creation and on being first-time-right. It developed an in-house self-service application CuVenSa that brought the various processes related to the creation, extension, change and deletion of business partner data into one platform. This application harmonized the previously segregated processes allowing data to have only a single point of entry and become a single source of truth. Moreover, the manual data check workflow was replaced by external data connections that automatically checked whether right data entered the system. This made data high-speed and correct. Subsequently, FashionC addressed its root cause by identifying exactly what went wrong with which data and where in the information supply chain. For this, it established an automated feedback loop starting

from the data consumers towards the data producers – connecting all relevant stakeholders under one chain of quality checks that quickly identifies data quality issues and immediately feeds back that information for the appropriate parties to fix. Improvement approaches depended on the type of root cause. For instance, data capture issues were resolved through data instance correction in the data attributes and usage error were treated by making data consumers aware regarding right information usage.

5.5 Control Phase

In this phase, ChemicalO aimed to communicate the success of the high-quality business partner data through dashboards exhibiting process-related KPIs such as 'process lead time'. The improvements were communicated to the wider organization because the data lifecycle processes supported many global functions. This built transparency and the opportunity for the data consumers to provide feedback to the data quality team in terms of upcoming use-cases that may need extra attention. FashionC adopted a more focused approach by developing stakeholder-specific data quality dashboards that communicate to only those relevant people to whom a particular data quality issue matter. This achieves transparency with only them who are concerned with the use-case in question. Such focused approach can make controlling more efficient and foster reduced lead time between issue identification and resolution.

6 DISCUSSION

The cases demonstrate that data quality initiatives can commence from either the input or output end in the information supply chain. These findings extend the pre-dominant approach of measuring mainly master data quality (Otto et al., 2010) and managing master data lifecycle (Ofner, Otto, Oesterle, & Straub, 2013) towards focusing on new and upcoming analytics use-cases driven by the increased usage and repurposing of data. Moreover, prioritization of data quality issues appeared to be a key action within the case studies. ChemicalO prioritized only on business partner data whereas FashionC concentrated on high-priority use-cases, showing that successful data quality improvements must be purposeful and cost-effective (Kleindienst, 2017) and a smart way to do so is to fix only what matters. We further observe that both firms placed high importance on creating visibility on data quality to create transparency and gain support. This supports literature that have argued for creating data quality awareness among stakeholders (McGilvray, 2021) and facilitating active participation in data quality activities using new methods (Zhang, Indulska, & Sadiq, 2019).We contribute to the extant literature which looked into data quality in only master data (Hikmawati et al., 2021), in information or data products (Machado, Costa, & Santos, 2021) or in enterprise systems (Glowalla & Sunyaev, 2014) towards an end-to-end view. We argue that a continuous monitoring and improvement cycle connects the relevant entities that play a key role in impacting data quality in the information supply chain. This paper also contributes to the practitioner knowledge by outlining an implementation blueprint regarding end-to-end data quality approaches. Our study comes with certain limitation. We studied only two organizations, missing other interesting data quality approaches with an end-to-end perspective. For future research, the concept and definition of data quality and relevant activities within the information supply chain should be further refined. This could provide basis for conceptualizing end-to-end data quality not only from the source and information product side, but also within the different data processing steps. Upcoming studies can also investigate end-to-end data quality approaches in tech-savvy companies versus in traditional ones.

REFERENCES

Barnaghi, P., & Sheth, A. (2016). On Searching the Internet of Things: Requirements and Challenges. *IEEE Intelligent Systems*, *31*, 71–75.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, *41*, 1–52.

Batini, C., Rula, A., Scannapieco, M., & Viscusi, G. (2015). From Data Quality to Big Data Quality. *Journal of Database Management (JDM)*, *26*, 60–82.

Batini, C., & Scannapieco, M. (2016). *Data and Information Quality: Dimensions, Principles and Techniques*. Cham: Springer International Publishing.

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The case research strategy in studies of information systems. *MIS Quarterly*, *11*, 369–386.

BusinessWire. (2020, September 1). New Industry Research Shows the Volume and Value of Data Increasing Exponentially in the Data Age. Retrieved January 11, 2022, from https://www.businesswire.com/news/home/20200901005035/en/New-Industry-Research-Shows-the-Volume-and-Value-of-Data-Increasing-Exponentially-in-the-Data-Age

Byabazaire, J., O'Hare, G., & Delaney, D. (2020). Data Quality and Trust: Review of Challenges and

Opportunities for Data Sharing in IoT. *Electronics*, *9*, 2083.

de Mast, J., & Lokkerbol, J. (2012). An analysis of the Six Sigma DMAIC method from the perspective of problem solving. *International Journal of Production Economics*, *139*, 604–614.

Eisenhardt, K. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, *14*, 532–550.

English, L. P. (1999). *Improving data warehouse and business information quality: Methods for reducing costs and increasing profits*. New York: Wiley.

Eppler, M. J., & Muenzenmayer, P. (2002). Measuring Information Quality In The Web Context: A Survey Of State-Of-The-Art Instruments And An Application Methodology. *Proceedings of the Seventh International Conference on Information Quality (ICIQ-02)*.

Glowalla, P., & Sunyaev, A. (2014). ERP system fit – an explorative task and data quality perspective. *Journal of Enterprise Information Management*, *27*, 668–686.

Haug, A., & Arlbjørn, J. S. (2011). Barriers to master data quality. *Journal of Enterprise Information Management*, 288–303.

Hikmawati, S., Santosa, P. I., & Hidayah, I. (2021). Improving Data Quality and Data Governance Using Master Data Management: A Review. *IJITEE (International Journal of Information Technology and Electrical Engineering)*, *5*, 90.

Jones-Farmer, L. A., Ezell, J. D., & Hazen, B. T. (2014). Applying Control Chart Methods to Enhance Data Quality. *Technometrics*, *56*, 29–41.

Klein, A., & Lehner, W. (2009). Representing Data Quality in Sensor Data Streaming Environments. *Journal of Data and Information Quality*, *1*, 1–28.

Kleindienst, D. (2017). The data quality improvement plan: Deciding on choice and sequence of data quality improvements. *Electronic Markets*, *27*, 387–398.

Krishnan, S., Haas, D., Franklin, M. J., & Wu, E. (2016). Towards reliable interactive data cleaning: A user survey and recommendations. *Proceedings of the Workshop on Human-In-the-Loop Data Analytics - HILDA '16*, 1–5. San Francisco, California: ACM Press.

Loshin, D. (2010). *Master Data Management*. Morgan Kaufmann.

Machado, I., Costa, C., & Santos, M. Y. (2021). *Data-Driven Information Systems: The Data Mesh Paradigm Shift*. 6. Valencia Spain: AIS.

McGilvray, D. (2021). *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information* (2nd edition). Waltham: Academic Press.

Metzger, A., Chi, C., Engel, Y., & Marconi, A. (2012). *Research Challenges on Online Service Quality Prediction for Proactive Adaptation*.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. SAGE.

Montgomery, D. C., & Woodall, W. H. (2008). An Overview of Six Sigma. *International Statistical Review / Revue Internationale de Statistique*, *76*, 329–346.

Ofner, M., Otto, B., Oesterle, H., & Straub, K. (2013). Management of the master data lifecycle: A framework for analysis. *Journal of Enterprise Information Management*, *26*, 472–491.

Otto, B., Ebner, V., & Hüner, Kai. M. (2010). *Measuring Master Data Quality: Findings from a Case Study*. Retrieved from https://core.ac.uk/reader/301348620

Otto, B., & Österle, H. (2015). *Corporate Data Quality Prerequisite for Successful Business Models*. Retrieved from http://nbn-resolving.de/urn:nbn:de:101:1-2015112720186

Paré, G. (2004). Investigating Information Systems with Positivist Case Research. *Communications of the Association for Information Systems*, *13*. https://doi.org/10.17705/1CAIS.01318

Parssian, A., Sarkar, S., & Jacob, V. S. (2004). Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product. *Management Science*, *50*, 967–982.

Röthlin, M. (2010). *Management of Data Quality in Enterprise Resource Planning Systems*. BoD – Books on Demand.

Si, Y., Xiao, Q., Su, J., Zeng, S., & Hong, X. (2020). Research on Data Product Quality Evaluation Model Based on AHP and TOPSIS. *Proceedings of the 4th International Conference on Computer Science and Application Engineering*, 1–5. Sanya China: ACM.

Singh, R., & Singh, D. K. (2010). A Descriptive Classification of Causes of Data Quality Problems in Data Warehousing. *International Journal of Computer Science Issues*, *7*, 10.

Smętkowska, M., & Mrugalska, B. (2018). Using Six Sigma DMAIC to Improve the Quality of the Production Process: A Case Study. *Procedia - Social and Behavioral Sciences*, *238*, 590–596.

Su, Z., & Jin, Z. (2007). A Methodology for Information Quality Assessment in the Designing and Manufacturing Processes of Mechanical Products [Chapter].

Taleb, I., Serhani, M. A., & Dssouli, R. (2018). Big Data Quality: A Survey. *2018 IEEE International Congress on Big Data (BigData Congress)*, 166–173. San Francisco, CA, USA: IEEE.

Tayi, G. K., & Ballou, D. P. (1998). Examining data quality. *Communications of the ACM*, *41*, 54–57.

Wang, R. Y. (1998). *A product perspective of TDQM. 41*.

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

Yin, R. K. (2003). *Case Study Research—Design and Methods* (3rd ed.). Sage Publications.

Zhang, R., Indulska, M., & Sadiq, S. (2019). Discovering Data Quality Problems: The Case of Repurposed Data. *Business & Information Systems Engineering*, *61*, 575–593.

Zhu, H., Madnick, S., Lee, Y., & Wang, R. (2014). Data and Information Quality Research: Its Evolution and Future. In H. Topi & A. Tucker (Eds.), *Computing Handbook, Third Edition* (pp. 16-1-16–20). Chapman and Hall/CRC.