UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

Critical Success Factors for Data Quality Management

ANA MARIA MARQUES RIBEIRO DOS SANTOS LUCAS

Orientadores:

Professor Doutor António Maria Palma dos Reis

Professor Doutor Mário Fernando Maciel Caldeira

Tese especialmente elaborada para obtenção do grau de Doutor em Gestão

2025

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR DE ECONOMIA E GESTÃO

Critical Success Factors for Data Quality Management

ANA MARIA MARQUES RIBEIRO DOS SANTOS LUCAS

Supervisors:

Professor Doutor António Maria Palma dos Reis

Professor Doutor Mário Fernando Maciel Caldeira

Tese especialmente elaborada para obtenção do grau de Doutor em Gestão

2025

Júri:

Presidente: Doutora Maria Rosa Vidigal Tavares da Cruz Quartin Borges, Professora Catedrática e Presidente do Conselho Científico do Instituto Superior de Economia e Gestão da Universidade de Lisboa.

Vogais:

Doutor Mário José Batista Romão, Professor Associado com Agregação do Instituto Superior de Economia e Gestão da Universidade de Lisboa;
Doutor Paulo Jorge Machado Oliveira, Professor Adjunto do Instituto Superior de Engenharia do Politécnico do Porto;
Doutor Fernando Brito e Abreu, Professor Associado, ISCTE - Instituto Universitário de Lisboa;
Doutor Jesualdo Cerqueira Fernandes, Professor Auxiliar do Instituto Superior de Economia e Gestão da Universidade de Lisboa;
Doutor António Maria Palma dos Reis, Professor Catedrático, do Instituto Superior de Economia e Gestão da Universidade Lisboa.

Acknowledgements

I would like to thank the many people without whose help this work would not have been possible.

First, I would like to thank my supervisors Professor António Palma dos Reis and Professor Mário Caldeira. Thank you for agreeing to supervise this thesis and thank you very much for being my friends.

My thanks also go to ISEG – Lisbon School of Economics and Management, and particularly the Management Department and all my colleagues, who made me feel at home.

To Professor Jesualdo Cerqueira Fernandes, for his friendship and for having been a kind of computer coach during the writing of this work.

To all the anonymous people who took part in the focus group, answered the Delphi questionnaire and collaborated on the case studies.

To Professor Ann Henshall for having the professionalism and patience to proofread my English.

Finally, but not least, to my sons, Frederico and Francisco, for all their love and support.

Abstract

Data and Information[1] are now widely regarded as an organization's most valuable assets, so they must be of high quality to maintain its competitive edge and support operational and decision-making processes. Today data quality is considered the biggest challenge to achieving success with generative artificial intelligence (GenAI)[2] currently very attractive to organizations.

Yet, although it is important for organizations to identify and prioritize critical success factors (CSF) for data quality management, the subject is still understudied. This work aims to help organizations to enhance their data quality by offering a set of Critical Success Factors (CSFs) and prioritizing their implementation based on their importance.

In line with Rockart (1979) we define critical success factors for data quality management as the limited number of areas in which the results, if satisfactory, guarantee data with better quality.

Currently, data quality research is conducted in two main scientific fields: computer science (CS) and management information systems (MIS) (Madnick et al., 2009). Our research falls specifically into the latter category.

Due to the nature and objectives of this study, it was decided to use the Design Science Framework, and, as an in-depth study was intended, qualitative methods were adopted. In the exploratory phase a focus group and a Delphi study were carried out, and in the explanatory one, two case studies.

This work identified two new critical success factors for data quality management, namely the creation of an information catalog and the use of data quality tools. In addition,

[1] Although data and information mean slightly different things, for reasons of simplicity, and in line with other data quality researchers, we use the terms data and information interchangeably.

[2] "Generative artificial intelligence is a type of artificial Intelligence (AI) technology that can build content such as audio, images, text, and videos. It involves algorithms such as ChatGPT, a chatbot that can produce essays, poetry, and other content requested by a user, and DALL-E, which generates art" (Campbell, 2023).

it developed a framework for the CSFs, organized into three clusters that represent the priorities for their implementation.

The order of importance of the CSFs did not coincide exactly between the Delphi study and the case studies, or between the two case studies, which may have happened because that the studies were carried out in different industries and the Delphi respondents belonged to multiple organizations from different sectors. Despite this, the most important CSFs almost coincide. In short, from this study it can be concluded that the nine CSFs in cluster A are the most reliable for use by organizations.

Rockart (1979) states that in general terms the CSFs differ between organizations, even in the same industry, and Xu & Lu (2003) point out that companies in different sectors do not attribute the same importance to some CSFs for data quality management.

Finally, it was not possible to identify the role of data governance in improving data quality management, namely whether it is a critical success factor or probably a core capability.

Keywords: critical success factors, data quality, data quality management, qualitative research, design science, focus group, Delphi study, case study

Resumo

Os dados e a informação[3] são atualmente considerados como os ativos mais valiosos das organizações, pelo que devem ser de elevada qualidade para manter a sua vantagem competitiva e apoiar os processos operacionais e de tomada de decisão. Para além disso a qualidade dos dados é hoje considerada como o maior desafio para alcançar o sucesso com a inteligência artificial generativa, atualmente muito procurada pelas organizações.

Para melhorar a qualidade dos seus dados importa que as organizações identifiquem e priorizem a utilização dos fatores críticos de sucesso (FCS) para a gestão dessa qualidade, tema que ainda está pouco investigado, pelo que este trabalho tem como objetivo ajudar as organizações que pretendam obter uma melhor qualidade dos dados, fornecendo um conjunto de FCS, e respetiva priorização, para melhorar essa qualidade.

Em linha com Rockart (1979) definimos fatores críticos de sucesso para a gestão da qualidade dos dados como o número limitado de áreas em que os resultados, se forem satisfatórios, garantem dados com melhor qualidade.

Atualmente a investigação sobre qualidade de dados é conduzida em duas áreas científicas principais (Madnick et al., 2009): ciência da computação (CC) e sistemas de informação de gestão (SIG), sendo que este trabalho se enquadra especificamente nesta última categoria.

Dada a natureza e os objetivos deste estudo, optou-se por utilizar a *Design Science Framework*, e como se pretendia um estudo aprofundado, decidiu-se utilizar métodos qualitativos, nomeadamente um *focus group* e um estudo Delphi na fase exploratória e dois estudos de caso na fase explanatória.

Este trabalho identificou dois novos fatores críticos de sucesso para a gestão da qualidade dos dados, nomeadamente a criação de um catálogo de informação e a utilização de ferramentas de qualidade dos dados, e conseguiu criar uma estrutura para os FCS, organizada em três clusters que representam as prioridades da sua utilização.

---

[3] Embora dado e informação sejam conceitos diferentes, por razões de simplicidade, e em consonância com outros investigadores da qualidade dos dados, utilizamos os termos dados e informações indistintamente.

A ordem de importância dos FCS não coincidiu exatamente entre o estudo Delphi e os estudos de caso, nem entre os dois estudos de caso, o que pode dever-se ao facto de os estudos terem sido elaborados em diferentes indústrias e os respondentes do Delphi pertencerem a múltiplas organizações de diferentes setores de atividade. Apesar disso os FCS mais importantes praticamente coincidem. Resumindo, deste estudo pode concluir-se que os nove CSFs do cluster A são os mais fiáveis para utilização pelas organizações.

Rockart (1979) afirma genericamente que os FCS diferem entre organizações, mesmo da mesma indústria e Xu & Lu (2003) referem que empresas de diferentes sectores não atribuem a mesma importância a alguns FCS para a gestão da qualidade de dados.

Por último, não foi possível identificar o papel do governo dos dados na melhoria da gestão da qualidade dos dados, nomeadamente se se trata de um fator crítico de sucesso ou provavelmente de um recurso essencial.

Palavras-chave: fatores críticos de sucesso, qualidade dos dados, gestão da qualidade dos dados, investigação qualitativa, *design science*, *focus group*, estudo Delphi, estudo de caso

Table of Contents

List of Figures

List of Tables

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
BCBS239	Principles for Effective Risk Data Aggregation and Risk Reporting
BDS	Business Data Steward
BP	IT Business Partner for Marketing and Sales
CCO	Chief Operations Center
CCS	Chief of Customer Service
CDO	Chief Data Officer
CEO	Chief Executive Officer
CIO	Chief Information Officer
CS	Computer Science
CSF	Critical Success Factors
DAMA	The Data Management Association
DAMA NL	The Data Management Association – The Netherlands
DB	Database
DC	Data Custodian
DG	Data Governance
DO	Data Owner
COC	Chief of Operations Center
DQ	Data Quality
DQC	Data Quality Champion
DQ-KPI	Data Quality Key Performance Indicator
DQM	Data Quality Management
DS	Data Steward
DSi	Data Scientist
EFQM	The European Foundation for Quality Management

ETL	Extract, Transform and Load
ID	Idiographic Design
IS	Information System
ISc	Idiographic Science
ISO	International Standards Organization
IT	Information Technology
KPI	Key Performance Indicators
MB	MyBank
MDM	Master Data Management
MIS	Management Information Systems
MIT	Massachusetts Institute of Technology
ND	Nomothetic Design
NS	Nomothetic Science
OC	Operations Center
OF	OwnFly
OS	Owner and Sponsor of Customer Master Data
PL	MDM Project Leader
PO	Process Owner
RDQ	Responsible for Data Management and Quality
RW	Real-World System
TDQM	Total Data Quality Management
TDS	Technical Data Steward
TDWI	Transforming Data with Intelligence
TQM	Total Quality Management

1 INTRODUCTION

1.1 Research Rationale

Academic and professional communities are increasingly concerned with data quality management. Today, there is a great deal of concern about the quality of corporate data since low quality data means inaccurate information, which can lead to resource waste and poor decision-making, harming the company, especially in terms of regulatory compliance and customer relationship management (Lucas, 2010, p. 162). Simultaneously, digital enterprise transformation and the growing use of big data, especially for predictive analysis, have resulted in an explosion of data in almost every industry and business sector, underscoring the value of data quality (Gao et al., 2016; Jin et al., 2015). The concept of "big data" has been around since the 1990s, with John Mashey being credited with popularizing it (Mashey, 1999). Following that, Laney (2001) defined the "Three V's" of Big Data (Volume, Variety and Velocity), with Volume meaning the large amounts of data to be managed, Variety the various types of data, more specifically structured, semi-structured and unstructured, and Velocity the speed of data generation and analysis. Since then, some more “V’s” have been added to the characteristics of Big Data; however, they are less consensual than the three initial V’s. They include Veracity, a term coined by IBM, which means the data’s quality, and Value, introduced by Oracle, which can be defined as the added value that the collected data can bring to a decision-making process (Kaur & Sood, 2017).

Kitchin (2013) defines the following characteristics of big data:

- Huge in size, including terabytes or petabytes of data;

- Quick in speed, emerging in real time or near real time;

- Diverse in scope, encompassing structured, semi-structured and unstructured data types;

- Broad in scope, seeking to encompass entire systems or populations;

- Fine-grained in resolution, aiming to be as detailed as possible;

- Flexible, extensible (can add new fields easily) and scalable (can expand in size rapidly).

Today data quality is considered the biggest challenge to achieving success with generative artificial intelligence (GenAI) (Addagada, 2024; Davenport & Tiwari, 2024), which is currently very attractive to organizations. According to a survey conducted by

Anaconda preparing and cleaning data accounts for 37.75% of data scientists' jobs (Anaconda, 2022). The job “data scientist” was coined in 2008 by D.J. Patil, and Jeff Hammerbacher, who defined it as “high-ranking professionals with the training and curiosity to make discoveries in the world of big data” (Davenport & Patil, 2012, p. 72). Furthermore, according to a survey of more than 3000 corporate executives (LaValle et al., 2011), cited by Hazen et al. (2014, p. 73), one in five executives believe that poor data quality is the main barrier to implementing more effective data analytic-based initiatives.

In 2016, IBM reported that the annual cost of bad data quality in the United States was $3.1 trillion (Redman, 2016). Similarly, according to a Royal Mail report published in 2016, low quality customer data cost UK businesses on average 5.9% of their annual sales (Royal Mail & DataIQ, 2016).

English (1999) divided the costs of bad data quality of customer addresses into three categories:

- Costs related to incorrectly sent or undelivered mail because of incorrect mailing addresses, which occur when procedures are not correctly carried out because of low-quality data;
- Information scraping and rework, which includes expenses for resending mail or for scraping and processing inaccurate data to meet the necessary quality standards; and
- Opportunity costs, which result from missed or lost sales opportunities. For instance, due to incomplete client address data for "loyalty cards", some cardholders do not appear in promotional offers, which can lower revenue.

Gartner reports that “poor data quality costs organizations an average $12.9 million per year” (Sakpal, 2021), and Loshin (2011) categorizes the impacts of poor data quality as follows:

- <u>Financial impacts</u> include higher operating expenses, lower income, lost opportunities, as well as fines, penalties, and other charges;
- <u>Implications related to confidence and satisfaction</u> include lower organizational trust, a lack of confidence in predicting, inconsistent operational and management reporting, and incorrect or delayed decision-making;

- Increased workloads, lower throughput, longer processing times, or lower-quality final products are examples of productivity consequences; and
- Impacts of risk and compliance include those related to credit evaluation, investment risks, fraud, as well as compliance with legal requirements or industry standards.

Although data and information mean slightly different things, for reasons of simplicity, and in line with other data quality researchers (Khatri & Brown, 2010; Madnick et al., 2009; Pipino et al., 2002), we use the terms data and information interchangeably.

Beginning in the late 1980s, data quality analysis mainly focused on computer science issues, such as developing techniques for querying different data sources and building large data warehouses (Madnick et al., 2009). In the early 1990s, MIT launched its Total Data Quality Management (TDQM) program and introduced the TDQM system, which extends to the domain of data the Total Quality Management (TQM) framework for quality improvement in the manufacturing domain (Madnick et al., 2009).

Currently, data quality research is conducted in two main scientific fields: computer science (CS) and management information systems (MIS) (Madnick et al., 2009). Our research falls specifically into the latter category.

1.2 Research Objective and Research Questions

The primary goal of this study is to identify the critical success factors for data quality management, as well as a priority for their implementation according to their importance. The analysis excludes the quality of big data, due to its moderate maturity and lack of clarity in its definition, a result of it being under-researched.

The research questions can be summarized as follows:

- What are the critical success factors (CSFs) for data quality management?
- What are the priorities for implementing the CSFs according to their importance?

Due to the nature of this research project, which seeks to create a new artifact (a critical success factor framework for data quality management), we use the design science framework, as explained in Chapter 3 (Gregor & Hevner, 2013; Hevner, 2007; Hevner et al., 2004; Iivari, 2007; Österle et al., 2011).

1.3 Thesis Structure

This thesis is organized into six chapters. This first chapter presents the research rationale, research objectives and research questions. Chapter two contains the literature review of data quality management (DQM) and critical success factors for DQM. The third chapter presents the philosophical perspective of the research and the choice of research design, taking into consideration the nature of the research objective and questions.

Chapter four presents the fieldwork, consisting of an exploratory phase followed by an explanatory one. In the first phase, a focus group was used, followed by a Delphi study, and in the second, two case studies were carried out, interpreted, and analyzed. Chapter five cross checks the findings of the Delphi study and those of the case studies and compares them with previous literature on the subject.

Chapter six proposes some research conclusions and presents the thesis contribution for the literature and practitioners. Lastly, we outline the study's limitations and suggest potential directions for further research.

2 LITERATURE REVIEW

This section reviews the literature on Data Quality Management (DQM) and Critical Success Factors for DQM.

2.1 Data, Information, and Knowledge

Hoffer et al. (2016, p. 41) define *data* as “stored representations of objects and events that have meaning and importance in the user’s environment” and *information* as “data that have been processed in such a way as to increase the knowledge of the person who uses the data”. English (1999) similarly presents information as a function f (Data, Metadata, Presentation) thus, data quality is determined by correct data values, consistent metadata definitions, and understandable presentation. *Knowledge,* according to Davenport & Prusak (1998, p. 4), is “a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. It originates and is applied “in the minds of knowers”. Peter Drucker (Drucker, 1989, p. 251) defines knowledge as “information that changes something or somebody, either by becoming grounds for actions, or by making an individual (or an institution) capable of different or more effective action”.

In line with Lucas (2010, p. 162), “data quality is ultimately intended to increase the productivity of the *knowledge worker* to create value for the business, as well as to assure data risk management and compliance”. The term "knowledge worker" was coined by Peter Drucker in 1959, and while there are various definitions available, they typically vary only in minor details. Knowledge workers, according to Sveiby (1997) are highly trained and educated professionals whose job primarily consists of using their competencies to transform information into knowledge, often with the help of suppliers of information or specialized knowledge. According to Drucker (2003, p. 169) “the most valuable asset of a twenty-first-century institution, whether business or non-business, will be its knowledge workers and their productivity”, and although the knowledge worker’s productivity depends on multiple factors, one is certainly related to the quality of data and information available.

2.2 The Data Quality Concept

Data quality (DQ) can be defined as “fitness for use by data consumers” (Wang & Strong, 1993, p. 6). This definition takes the consumer viewpoint, in line with the general literature on quality. Some of the more general definitions of data quality, all of which consider, directly or indirectly, the point of view of data consumers, are presented in Table I. The related concept of "data quality dimension" can be defined as a “set of data quality attributes that represent a single aspect or concept of data quality” (Wang & Strong, 1993, p. 6). Examples of data quality dimensions are accuracy, timeliness, completeness, and consistency. High-quality data should be “intrinsically good, contextually appropriate for the task, clearly represented, and accessible to the data consumer” (Wang & Strong, 1993, p. 6).

Table I – Data Quality Definitions

Data quality is data that are fit for use by data consumers	Wang & Strong (1993, p. 6)
Data are of high quality if they are fit for their intended uses in operations, decision making, and planning	Juran & Godfrey (1999, p. 998)
Data quality is the degree to which data is accurate, complete, timely, consistent with all requirements and business rules, and relevant for a given use	Mosley et al. (2008, p. 42)
Data quality is the "degree to which a set of inherent characteristics of data fulfills requirements"	(ISO Standard 8000 Part (2) Version (1) - Data Quality?: Vocabulary, 2017, p. 4)
Data quality is the "degree to which the characteristics of data satisfy stated and implied needs when used under specified conditions"	(ISO/IEC 25012, 2008, p. 3)

Universal (domain-independent) dimensions for data quality have been identified and described by means of three main scientific approaches and two ISO standards:

- Theoretical (Wand & Wang, 1996);
- Empirical (Wang & Strong, 1993);
- Intuitive (Redman, 1996);
- ISO/IEC (ISO/IEC 25012, 2008);
- ISO 8000 (ISO 8000-8:2015, 2015).

These approaches refer to both the data in extension, i.e., their values (English, 1999; ISO 8000-8:2015, 2015; ISO/IEC 25012, 2008; Redman, 1996; Wand & Wang, 1996; Wang & Strong, 1993) and in intention, i.e., their models or database schemas and business rules that apply to the data (English, 1999; ISO 8000-8, 2015; Redman, 1996). Some authors (English, 1999; ISO 8000-8:2015, 2015; ISO/IEC 25012, 2008; Redman, 1996; Wang & Strong, 1993) also consider the following to be dimensions: data presentation and contextual data quality, as well as system-dependent dimensions, for instance, data security and accessibility dimensions (ISO 8000-8:2015, 2015; ISO/IEC 25012, 2008; Wang & Strong, 1993).

Regarding data model dimensions, both the professional and research literatures only consider intra-data model dimensions (those related to the conceptual and logical model of a specific database), eschewing inter-data model dimensions, which apply to the representation of the same entity in different databases. Oliveira et al. (2005) present a detailed description of data quality problems that can appear when the same data exists in different databases.

Although there are other approaches, they are less recognized than the five mentioned above and add very little to those analyzed in this document.

2.3 The Theoretical Approach

Wand & Wang (1996) base their analysis on ontological constructs and provide a theoretical approach to the information system "internal view" data quality dimensions. They make a distinction between the external and internal views of an information system, considering that the external view is focused on how an information system is used and what its impact is. By contrast, the internal view considers the requirements which reflect the external view and addresses the construction and operation of the information system to achieve the required functionality.

In this approach the information system (IS) is viewed as a representation of a real-world system (RW), and RW is said to be properly represented if: "(1) an exhaustive mapping, RW → IS, exists, and (2) no two states in RW are translated into the same state in IS (the inverse mapping is a function") (Wand & Wang, 1996, p. 91)(Figure 1).

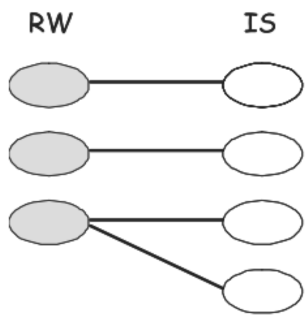

Figure 1 - Proper representation of the real-world system

Source: (Wand & Wang, 1996, p. 90)

Data shortcomings result from any divergence from proper representation, which can be caused by system design or system operation flaws. There are three types of design flaws: "incomplete representation," "ambiguous representation," and "meaningless states (Wand & Wang, 1996, p. 91) (Figure 2).

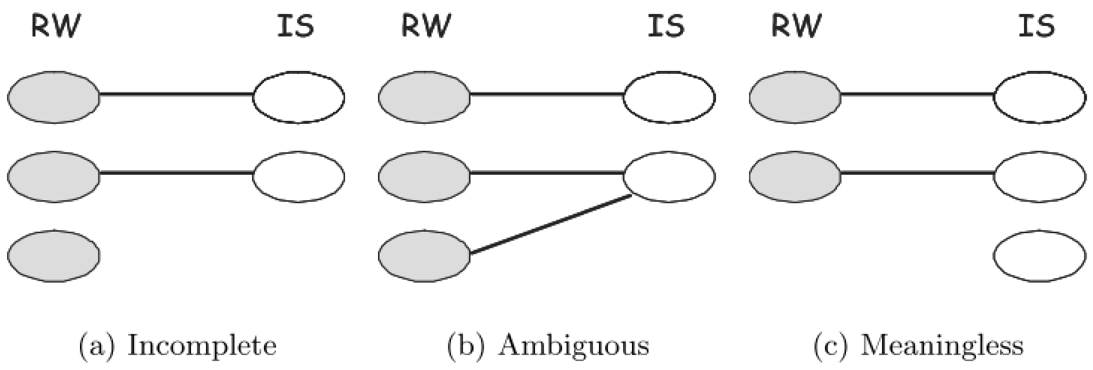

Figure 2 - Incomplete, Ambiguous and Meaningless representations of the real world

Source: (Wand & Wang, 1996, p. 90)

Furthermore, a state in RW could be mapped to an erroneous state in IS during operation, a scenario known as garbling. The authors distinguish between two scenarios: mapping to a meaningless state (assuming meaningless states exist in the IS), or mapping to a meaningful but inaccurate IS state (Figure 3).

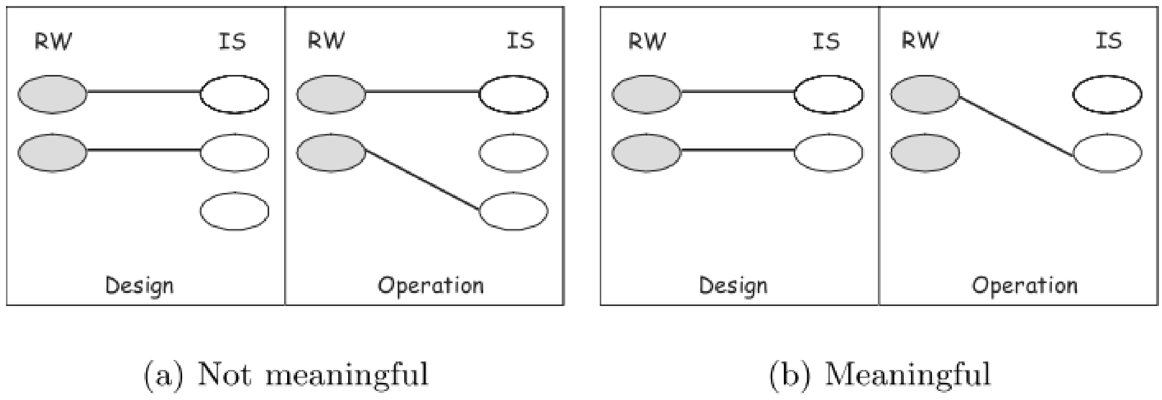

Figure 3 - Garbling representations of the real-world system

Source: (Wand & Wang, 1996, p. 91)

Using the literature and referring to the above-mentioned flaws, a set of data quality dimensions for the internal view (Wand & Wang, 1996, p. 93) is defined:

Accuracy – “Inaccuracy indicates that IS represents a different RW state than the one that should have been represented”;

Reliability – “Measure of agreement between expectations and capability, and as how data conforms with user requirements or reality”;

Timeliness – “The delay between a change in the RW state and the consequent modification of the IS state. Due to a lack of timeliness, the IS may reflect the RW previous state”;

Completeness – “An IS ability to represent any meaningful state of the depicted RW system”;

Consistency – If more than one state of the IS matches a state of the RW system, data values must be consistent.

It is considered that *timeliness* dimension belongs simultaneously to the internal and external views, and the remaining dimensions belong only to the internal view.

2.4 The Empirical Approach

Wang & Strong (1993) propose 15 data quality dimensions, which were developed through the analysis of two surveys to data consumers. The first survey intended to compile a comprehensive list of potential data quality dimensions, while the second assessed their relative values. The selected data quality dimensions are organized into four categories*: intrinsic, contextual, representational, and accessibility* (Figure 4).

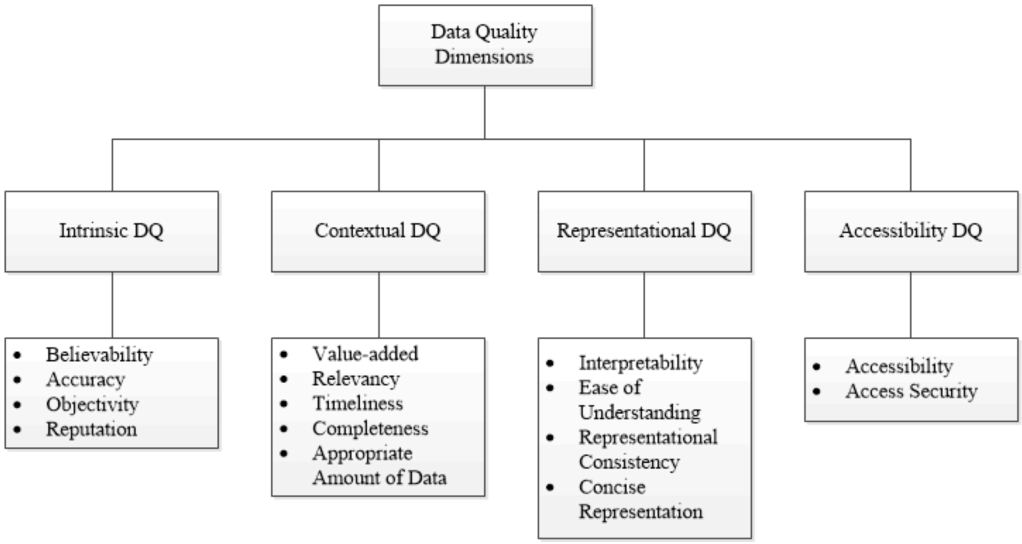

Figure 4 - A Conceptual Framework for Data Quality Dimensions

Adapted from (Wang & Strong, 1993, p. 20)

The categories are defined as follows (Wang & Strong, 1993):

- *Intrinsic DQ* - The degree to which data values are consistent with their true or actual values;
- *Contextual DQ* – The extent to which data relate to or apply to the data user's task;
- *Representational DQ* – The degree to which data are presented in a clear and understandable manner;
- *Accessibility DQ* – The degree to which data are available or obtainable.

Table II presents the definitions of dimensions (Kahn et al., 2002).

Table II - Data Quality Dimensions

Dimensions	**Definitions**
Believability	The degree to which data are considered as true, real, and credible
Accuracy	The degree to which data are accurate, reliable, and error-free
Objectivity	The degree to which data are unbiased and impartial
Reputation	The extent to which data may be trusted or highly regarded in terms of their source or content,
Value-Added	The extent to which data are useful and produce benefits because of their use
Relevancy	The extent to which data are relevant and useful for the task at hand
Timeliness	The extent to which the data's age is acceptable for the task at hand
Completeness	The extent to which the data are sufficient in terms of breadth, depth, and scope for the task at hand
Appropriate Amount of Data	The extent to which the quantity or volume of data available is appropriate
Interpretability	The extent to which data are in acceptable language and units, and data definitions are clear
Ease of Understanding	The degree to which data are unambiguous and simple to understand
Representational Consistency	The degree to which data are always presented in the same manner and are interchangeable with previous data
Concise Representation	The degree to which data are compactly conveyed without becoming overwhelming (i.e., brief in presentation, yet complete and to the point)
Accessibility	The extent to which data are readily available or retrievable
Access Security	The degree to which data access can be restricted and hence maintained secure

Source: (Kahn et al., 2002, p.187)

The data quality dimensions presented in Table II have now become largely accepted in the literature, for example in (Baskarada & Koronios, 2014; Madnick et al., 2009)

2.5 The Intuitive Approach

Redman (1996) defined data quality dimensions intuitively and organized them into three categories, namely *conceptual schema*, *data values*, and *data format*. These categories are presented respectively in Tables III, IV and V.

Table III- Data Quality Dimensions of the Conceptual Schema

Groups	Dimensions	Definitions
Content	Relevance	The schema should supply the data that the applications require
	Obtainability	Data values should be easily obtainable
	Clarity of Definition	Each attribute in the schema definition should be clearly defined
Scope	Comprehensiveness	Each needed attribute should be included
	Essentialness	There should be no unnecessary attributes included
Level of Detail	Attribute Granularity	To support applications, the attributes should be defined to the appropriate level of detail
	Domain Precision	The possible value domains should be just large enough to support applications
Composition	Naturalness	Each attribute in the schema should have a "natural" counterpart in the real world
	Identifiability	Individual entities should be straightforward to recognize thanks to the schema
	Homogeneity	Entity types should be defined to reduce the number of attributes that aren't needed
	Minimum Redundancy	Redundancy should be reduced to a bare minimum
Schema Consistency	Semantic Consistency	Consistency of the various components of the schema
	Structural Consistency	Wherever possible, entity types and attributes should have the same basic structure
Reaction to Change	Robustness	The schema should be sufficiently flexible such that it does not need to be changed every time an application change
	Flexibility	The schema can be altered easily if necessary

Adapted from Redman (1996, p. 247, 267)

Table IV - Data Quality Dimensions of Data Values

Dimensions	Definitions
Accuracy	The accuracy of a datum < e, a, v> refers to the value's proximity of a value *v* to some value *v'* in the attribute domain that is thought to be the proper one for entity *e* and attribute *a*
Consistency	When it comes to a set of constraints, data are consistent if they satisfy all of them
Completeness	The degree to which values are present in a data collection
Currency	The degree to which a datum is up to date

Adapted from Redman (1996, p. 254-263)

Table V - Data Quality Dimensions of Data Formats

Dimension	Definitions
Appropriateness	If one format is better suited to user needs than another, it is more appropriate
Interpretability	A suitable format aids the user in appropriately interpreting values
Portability	Good formats are portable or universal
Format precision	Users must be able to distinguish between domain elements
Format flexibility	Changes in user requirements and recording media can be adjusted easily
Ability to represent null values	It is possible to distinguish null and default values from applicable values of the domain
Efficient use of storage	In terms of physical representation, efficiency is key. A code is more efficient than an icon
Representation consistency	There is coherence between physical data objects and their formats

Adapted from Redman (1996, p. 260-263)

2.6 The ISO/IEC 25012 Approach

ISO/IEC 25012 (2008) classifies the dimensions of data quality into inherent data quality and system-dependent data quality. Using the term characteristics instead of dimensions, the standard defines *inherent data quality* as the degree to which quality characteristics of data have the inherent capacity to satisfy stated and implied needs. It defines *system-dependent data quality* as the degree to which data quality is kept and preserved within a computer system when data are used under specified conditions. Table VI presents the 15 data quality characteristics, and Table VII presents their definitions.

Table VI - Data Quality Model Characteristics

Characteristics	Data Quality	
	Inherent	System dependent
Accuracy	X	
Completeness	X	
Consistency	X	
Credibility	X	
Currentness	X	
Accessibility	X	X
Compliance	X	X
Confidentiality	X	X
Efficiency	X	X
Precision	X	X
Traceability	X	X
Understandability	X	X
Availability		X
Portability		X
Recoverability		X

Source: (ISO/IEC 25012, 2008, p. 5)

Table VII - Data Quality Characteristics

Characteristic	Definition
Accuracy	The degree to which data have attributes that correctly represent the true value of the intended attributes of a concept or event in a specific context of use. It has two main aspects: 1. **Syntactic accuracy** is defined as the closeness of the data values to a set of values defined in a domain considered syntactically correct; 2. **Semantic accuracy** is defined as the closeness of the data values to a set of values defined in a domain considered semantically correct.
Completeness	The degree to which subject data associated with an entity include values for all expected attributes and related entity instances in a specific context of use.
Consistency	The degree to which data have attributes that are free from contradiction and are coherent with other data in a specific context of use. It can be either or both among data regarding one entity and across similar data for comparable entities.
Credibility	The degree to which data have attributes that are regarded as true and believable by users in a specific context of use.

Source: (ISO/IEC 25012, 2008, pp. 6–10)

Table VII - Data Quality Characteristics (continued)

Characteristic	Definition
Currentness	The degree to which data have attributes that are of the right age in a specific context of use.
Accessibility	The degree to which data can be accessed in a specific context of use, particularly by people who need supporting technology or special configuration because of some disability.
Compliance	The degree to which data have attributes that adhere to standards, conventions or regulations in force and similar rules relating to data quality in a specific context of use.
Confidentiality	The degree to which data have attributes that ensure that they are only accessible and interpretable by authorized users in a specific context of use.
Efficiency	The degree to which data have attributes that can be processed and provide the expected levels of performance by using the appropriate amounts and types of resources in a specific context of use.
Precision	The degree to which data have attributes that are exact or that provide discrimination in a specific context of use.
Traceability	The degree to which data have attributes that provide an audit trail of access to the data and of any changes made to the data in a specific context of use.
Understandability	The degree to which data have attributes that enable them to be read and interpreted by users, and are expressed in appropriate languages, symbols and units in a specific context of use.
Availability	The degree to which data have attributes that enable them to be retrieved by authorized users and/or applications in a specific context of use.
Portability	The degree to which data have attributes that enable them to be installed, replaced or moved from one system to another preserving the existing quality in a specific context of use.
Recoverability	The degree to which data have attributes that enable them to maintain and preserve a specified level of operations and quality, even in the event of failure, in a specific context of use.

2.7 The ISO 8000-8 Approach

ISO 8000-8:2015 (2015) defines the following three categories for data quality:

- *Syntactic quality*, which is the extent to which the metadata's requirements—that is, the data's prescribed syntax—are met;
- *Semantic quality*, which is the extent to which data match what they are supposed to portray;

- *Pragmatic quality*, which is the extent to which data are deemed valuable and appropriate for a specific purpose.

The dimensions of the three categories for data quality are described below.

Table VIII – Syntactic Quality Category

Dimension	Definition
Entity Integrity	Every entity shall have a unique primary key
Referential integrity	Every entity that is referenced shall exist as an entity in its own right
Domain integrity	All attribute values shall be within the specified domain
User defined integrity	All user defined constraints shall be complied with

Source: (ISO 8000-8:2015, 2015, p. 7)

Table IX- Semantic Data Quality Category

Dimension	Definition
Mapped completely	Every entity inside the domain of interest must be represented
Mapped consistently	One of the following shall represent each entity in the domain of interest: - One identifiable data unit, at most; - Numerous identifiable units that are consistent; - Multiple identifiable units whose discrepancies are fixed in a reasonable amount of time
Mapped meaningfully	Each identifiable data unit shall represent at least one specific entity in the domain of interest
Mapped unambiguously	Each identifiable data unit shall represent at most one specific entity in the domain of interest
Entities in the domain of interest mapped correctly	Each identifiable data unit shall map to the correct entity in the domain of interest
Properties mapped correctly	All attribute values in an identifiable data unit shall match the property values for the represented entity in the domain of interest

Source: (ISO 8000-8:2015, 2015, p. 3-4)

Table X – Pragmatic Data Quality Category

Dimension	Definition
Accessible	Information is easy and quick to retrieve
Complete	In a trustworthy 1:1 mapping, information is thought to be fully mapped to entities in the domain of interest. Users' perceptions of semantic quality are included in this component
Flexible content	The information content can be altered as needed, including the measurement units, degree of precision, and level of detail. Every important piece of metadata is available
Flexible layout	The layout can be modified as needed. It is possible to specify the data to retrieve and how they are formatted
Secure	The information is suitably shielded against misuse or harm (including illegal access, use, or dissemination)
Useful	The information is useful. The information is understandable, and suitable for its purpose, including cost and benefit aspects. The currency (age) of the data is appropriate for their use

Source: (ISO 8000-8:2015, 2015, p. 8)

2.8 Comparative Analysis of Dimensions

According to the dimensions presented in the preceding sections, there is no agreement among the various authors on either the set of dimensions or the definitions of those with the same denomination. For instance, the definitions of timeliness in the theoretical and empirical approaches differ as explained below. There follows a synthesis of the different approaches, a brief comparison of the dimensions of time and consistency or coherence of data values, and a note on data quality in big data.

To compare the different approaches, we will borrow the categories of the empirical approach (Wang & Strong, 1993) and their definitions, namely Intrinsic DQ, Contextual DQ, Representational DQ and Accessibility DQ, renaming the latter System Dependent DQ for the sake of completeness. We also add a category Schema DQ, defined as the quality of the logical schema, to accommodate the "Conceptual Schema" category from the intuitive approach (Redman, 1996) and "Syntactic Quality" from ISO (ISO 8000-8:2015, 2015). Table XI below presents the mapping of approaches into the categories.

Table XI – Approaches vs Categories

Category / Approach	Intrinsic DQ	Contextual DQ	Representational DQ	System Dependent DQ	Schema DQ
Theoretical	Yes	Yes	No	No	No
Empirical	Yes	Yes	Yes	Yes	No
Intuitive	Yes	Yes	Yes	No	Yes
ISO/IEC 25012	Yes	Yes	Yes	Yes	No
ISO 8000-8	Yes	Yes	Yes	Yes	Yes

Source: the author

The results show that ISO 8000-8 is the only approach that satisfies all categories, and that the theoretical approach only satisfies Intrinsic DQ and Contextual DQ, because the authors assume they only consider the internal view of the system, that deals with the requirements which reflect the external view and addresses the construction and operation of the information system to achieve the required functionality. The only approaches that comply with Schema DQ category are the Intuitive and ISO 8000-8, and the Intuitive approach does not consider the System Dependent DQ category.

With regard to time dimensions, *timeliness* in the theoretical approach represents the same dimension as *currency* in the intuitive approach, representing the “degree to which a datum is up-to-date” (Redman, 1996) and *timeliness* in the empirical approach has the same meaning as *currentness* in the ISO/IEC 25012 and represents “the extent to which the age of the data is appropriate for the task at hand” (ISO/IEC 25012, 2008; Wang & Strong, 1993). Thus, data that are not up to date can have good timeliness if they are appropriate for the task at hand, such as data used in longitudinal studies.

Batini & Scannapieco (2006, p. 29) clearly define three time-related dimensions:

- Currency concerns how promptly data are updated;
- Volatility characterizes the frequency with which data vary in time;
- Timeliness expresses how current data are for the task at hand.

It should be noted that the standard ISO/IEC 25012 (2008) presents interesting dimensions to consider, namely compliance and traceability important, for example, for the bank's compatibility with the standard BCBS 239[4] (Principles for Effective Risk Data Aggregation and Risk Reporting, 2013), with which they must comply.

Although the consistency or coherence of data values dimension is defined in several studies (Batini & Scannapieco, 2006; ISO 8000-8:2015, 2015; ISO/IEC 25012, 2008; Redman, 1996; Wand & Wang, 1996), the definitions are always very general, without specifying whether the coherence is within one database or between several databases. Moreover, the consistency of the same data in different databases is never mentioned, a problem that has already been noted regarding inter-model or inter-schema dimensions. Notwithstanding this absence, definitions that include inter-model coherence in the dimension of coherence do appear in the literature. In a health report on patient registries Zaletel & Kralj (2015, p. 64) define the data quality dimension coherence (or consistency) as "coherence that covers the internal consistency of a data collection as well as its comparability both over time and with other data sources". Similarly, in a survey of the literature on data quality dimensions, DAMA The Netherlands (DAMA NL) Black & van Nederpelt (2020, p. 52) define coherence as "the degree to which datasets can be combined", but unfortunately, they do not provide a bibliographic reference that supports their definition.

Oliveira et al. (2005) present a detailed description of data quality problems that can appear when the same data exists in different databases.

One last consideration relates to the data quality dimensions of big data. The little literature available states that many of the dimensions of structured data apply to big data. Nevertheless, in some cases the definitions differ, and the redefinition of the dimensions depends on the type of data, such as maps, semi structured texts or linked open data (Batini et al., 2015; Gabr et al., 2021; Ramasamy & Chowdhury, 2020).

2.9 Data Quality Management and Data Governance

Data Quality Management (DQM) can be defined as the "quality-oriented management of data as an asset" (Weber et al., 2009, p 4:4), that is, the "the application of total quality management (TQM) concepts and practices to improve data and information quality,

---

[4] The purpose of these rules is to improve banks' risk management.

including setting data quality policies and guidelines, data quality measurement (including data quality auditing and certification), data quality analysis, data cleansing and correction, data quality process improvement, and data quality education" (DAMA, 2008, p. 43). According to (ISO Standard 8000 Part (2) Version (1) - Data Quality: Vocabulary, (2017, p. 4), data quality management is a set of "coordinated activities to direct and control an organization with regard to data quality".

To be effective, data quality management must go beyond repairing bad data to prevent data quality issues by managing data throughout their life cycle to meet the information needs of its stakeholders. Furthermore, to address both organizational and technological viewpoints, DQM must break down the stovepipes that segregate data across business units and foster collaboration between business and IT divisions. This approach requires a significant cultural shift, requiring leadership, authority, control, and resource allocation. As a result, it needs governance, particularly data governance (DG) (Lucas, 2010, p. 163), which can be defined as "the exercise of authority, control and shared decision making (planning, monitoring, and enforcement) over the management of data assets" (DAMA, 2017, p. 67).

Figure 5 - DAMA-DMBOK2 Data Management Framework

Source: (DAMA, 2017, p. 449)

Nevertheless, DG does not equal DQM (Figure 5), either regarding the entity that makes the decisions or to their scope. The board of directors and executive management are responsible for DG, which is primarily focused on the business environment and strategic directions, encompassing topics other than DQM, such as data security and privacy and information life-cycle management (IBM Data Governance Council, 2007). Organizations can use data governance to establish corporate-wide DQM accountability, involving professionals from both the business and IT departments (Wende, 2007).

Khatri & Brown (2010) identify five decision domains for Data Governance, one of which is data quality (Figure 6).

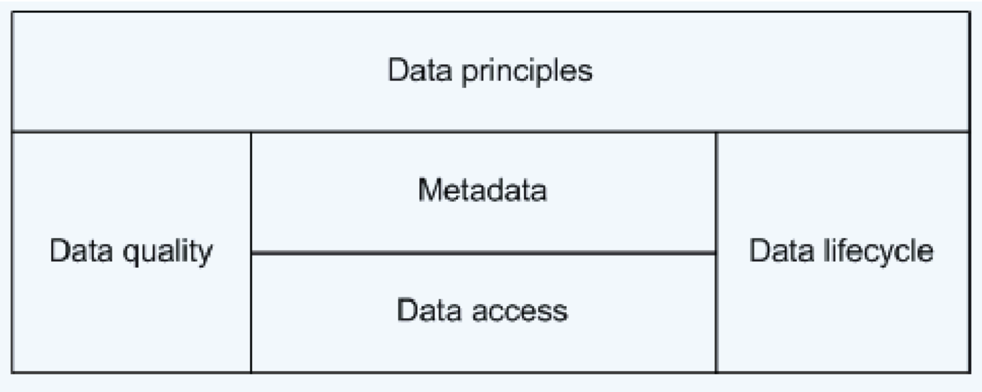

Figure 6 – Decision Domains for Data Governance
Source: (Khatri & Brown, 2010, p. 149)

2.10 Roles and responsibilities in the Data Quality Management Approach

Literature - whether from academic or professional sources - presents a set of roles and responsibilities along with their definitions related to the management of data, considered to be a corporate asset. Table XII presents a non-exhaustive set of roles and responsibilities within the DQM approach, which we are using in this study.

Table XII – DQM Roles

Role	Description
Data Owner (DO)	"An individual responsible for policy and practice decisions about data. For business data, the individual may be called a business owner of the data" (DAMA, 2008, p. 42).
Data Steward (DS)	According to DAMA (2008, p. 45), a "DS is a business leader and/or subject matter expert designated as accountable for: - The identification of operational and business intelligence data requirements within an assigned subject area; - The quality of data names, business definitions, and domain values within an assigned subject area; - Compliance with regulatory requirements and conformance to internal data policies and data standards; - Application of appropriate security controls; - Analysis and improvement of data quality; - Identification and solution of data-related issues." There are different types of Data Stewards, and we will present the most widely used: Business Data Stewards (BDS) and Technical Data Stewards (TDS). Business Data Stewards (BDS) – "A knowledge worker, business leader and recognized subject matter expert assigned accountability for the data specifications and data quality of specifically assigned business entities, subject areas or databases" (DAMA, 2008, p. 17) Technical Data Stewards (TDS) - TDS "provide support and are associated with specific systems, applications, data stores and technical processes such as Identity Resolution, data quality rule enforcement, and Extract, Transform and Load (ETL)" (Plotkin, 2021, p.32)
Data Custodian (DC)	"An IT individual or organization responsible for the IT infrastructure providing and protecting data in conformance with the policies and practices prescribed by data governance" (DAMA, 2008, p. 36)
Data Quality Champion (DQC)	According to Tee et al. (2007, p. 338), DQCs "are managers who actively and vigorously promote their vision for using data quality-related technology innovations. They push projects through approvals, provide political support, keep participants informed, and allocate resources to data quality projects"
Process Owner (PO)	PO has "end-to-end responsibility for critical information chains" (Redman, 1996, p. 274)

Table XII – DQM Roles (continued)

Role	Description
Chief Data Officer (CDO)	A Chief Data Officer (CDO) "assists in bridging the technological and business divide. At a senior level, this person promotes an enterprise-wide data management strategy. The CDO oversees data management activities, which enable a business to take advantage of its data assets and create a competitive advantage. A CDO is a mix of business strategist, consultant, data quality guardian, and data management evangelist" (Knight, 2021)

2.11 Critical Success Factors

The concept of Critical Success Factor (CSF) was introduced by D. Ronald Daniel in 1961 (Daniel, 1961) and further popularized by John F. Rockart (Rockart, 1979) According to Rockart (1979, p. 85),

> *Critical success factors are, for any business, the limited number of areas in which results, if they are satisfactory, will ensure successful competitive performance for the organization.*

In the context of this work and in line with Rockart, we define critical success factors for data quality management as the limited number of areas in which results, if they are satisfactory, will ensure data with better quality.

In his seminal work on Critical Success Factors, and based on case studies, Rockart (1979) states that CSFs differ among companies, even in the same industry, and even among managers. He, nevertheless, suggests that certain CSFs are "all-encompassing industry-based factors" and the rest "are generated by differences in environmental situation, temporal factors, geographic location, or strategic situation" (Rockart, 1979, p. 87).

Rockart (1979) categorized CSFs into two groups: monitoring and building. The CSFs in the former group serve, as the name implies, to monitor the current results, and the CSFs in the latter group are oriented towards building for the future.

2.12 Critical Success Factors for Data Quality Management

The literature on CSFs for data quality management is scarce, with previous studies dealing with their identification and rating (Akpon-Ebiyomare et al., 2012; Baskarada &

Koronios, 2014; Santos, 2015; Xu, 2015; Xu & Lu, 2003). Some papers rate CSFs in different industries (Xu & Lu, 2003) or different sized organizations (Xu, 2003), while others only rate the social, cultural, and organizational CSFs. Given this situation, it was decided to start the literature review by CSFs for Total Quality Management (TQM) because it is a much more mature subject. Furthermore, data quality management can be considered a sub-area of TQM, because there is an analogy between the quality aspects between the manufacturing of products and those of information (Wang, 1998).

The literature review on CSFs for TQM is based on a review of papers that are themselves literature reviews (Black & Porter, 1996; Hietschold et al., 2014; Porter & Parker, 1993; Saraph et al., 1989). It also covers the criteria underlying two important quality awards, one for the USA, Malcolm Baldrige National Quality Award (NIST, 2017) and the other for Europe, EFQM[5] Excellence Model (EFQM, 2013) and the ISO Quality Management Principles (ISO 9001:2015, 2015).

Table XIII summarizes the literature review on CSFs for TQM and Table XIV on CSFs for DQM. Table XV explains each CSF for DQM.

[5] The European Foundation for Quality Management

Table XIII - Critical Success Factors for Total Quality Management

Nb	Critical Success Factors	Saraph, Benson, Schroeder (1989)	Hietschold, Reinhardt, Gurtner (2014)	Porter &Parker (1993)	Black & Porter (1996)	EFQM Excelence Model (2013)	Malcolm Baldrige National Quality Award NIST(2017)	ISO Quality Management Principles ISO 9001:2015 (2015)
1	**Management commitment and leadership**	The role of management leadership and quality policy	Top management commitment and leadership	Management behavior	Strategic quality management	Leadership	Leadership/ Results (Leadership and governance results)	Leadership
2	**Organization for quality**	Role of quality department		Organization	Teamwork structures for improvement			
3	**Training and learning**	Training	Training and learning	Training and education	Communication of Improvement Information		Workforce	
4	**Product/Service Design**	Product/Service Design			External Interface Management		Operations	
5	**Supplier parnership**	Supplier quality management	Supplier parnership		Supplier partnerships	Partnership and resources		Relationship management
6	**Process management**	Process management	Process management	Process Management and systems	Quality Improvement Measurement Systems /Stategic Quality Management	Processes, products and services	Operations	Process approach
7	**Quality data and reporting**	Quality data and reporting	Information/analysis/data	Quality Technologies	Quality Improvement Measurement Systems / Strategic quality management		Measurement, analysis, and knowledge management	Evidence-based decision making
8	**Employee involvement**	Employee relations	HRM/recognition/teamwork	Employee involvement	People and Customer Management/Strategic quality management	People/People results	Workforce/ Results (Workforce-Focused Results)	Engagement of people
9	**Customer focus and satisfaction**		Customer focus and satisfaction		People and Customer Management/Customer Satisfaction Orientation/ Erternal Interface Management	Customer results	Customers/ Results: (Customer focused results)	Customer focus
10	**Strategic quality planning**		Strategic quality planning	Strategy	Strategic Quality Management	Strategy	Strategy	

Table XIII– Critical Success Factors for Total Quality Management (continued)

Nb	Critical Success Factors	Saraph, Benson, Schroeder (1989)	Hietschold, Reinhardt, Gurtner (2014)	Porter &Parker (1993)	Black & Porter (1996)	EFQM Excelence Model (2013)	Malcolm Baldrige National Quality Award NIST(2017)	ISO Quality Management Principles ISO 9001:2015 (2015)
11	**Operations**				Operational quality management		Operations/Results (Product and process results)	
12	**Culture and communication**		Culture and communication	Communication	Communication of Improvement Information/ Corporate quality culture	People		
13	**Benchmarking**		Benchmarking					
14	**Social and environmental responsibility**		Social and environmental responsibility		External Interface Management	Society results	Leadership (Governance and societal responsibilities)	
15	**Continuous improvement**							Improvement

In the construction of the CSF for the DQM list, we removed the CSF Benchmarking and Social and Environmental Responsibility because they are not referenced in the DQM literature (see Table XIV).

Table XIV - Critical Success Factors for Data Quality Management

Nb	Critical Success Factors	CSF for TQM	Baškarada & Koronios (2014)	Xu, Koronios & Brown (2003)	Xu (2015)	Akpon-Ebiyomare et al. (2012)
1	**Management commitment and leadership**	Management commitment and leadership	Information quality management governance	Top management / Middle management commitment to DQ	Top management/Middle management commitment to DQ	Management commitment & support
2	**Organization for quality**	Organization for quality		Organizational structure	Appropriate organizational structure/Establish DQ manager position	
3	**Training**	Training and learning	Training	Education and training	Education and training	Training and communication
4	**Information quality requirements management**	Product/Service Design	Information quality requirements management			
5	**Supplier parnership**	Supplier parnership		Data supplier quality management	Data supplier quality management	Information supplier quality management
6	**Information product lifecycle management**	Process management	Information product lifecycle management			
7	**Information quality assessment/monitoring**	Quality data and reporting	Information quality assessment/monitoring		Measurement and reporting	
8	**Teamwork**	Employee involvement		Teamwork	Teamwork	Teamwork between different departments and within departments
9	**Customer focus and satisfaction**	Customer focus and satisfaction	Information quality requirements management	Customer focus	User focus	Customer focus/User involvement
10	**Strategic Data Quality Planning**	Strategic quality planning				

Table XIV– Critical Success Factors to Data Quality Management (continued)

Nb	Critical Success Factors	CSF for TQM	Baškarada & Koronios (2014)	Xu, Koronios & Brown (2003)	Xu (2015)	Akpon-Ebiyomare et al. (2012)
11	**Continuuous information quality management improvement**	Operations	Continuuous information quality management improvement			
12	**Culture and communication**	Culture and communication	Organizational culture of focusing on DQ		Organizational culture of focusing on DQ/Clear DQ vision for entire organization	Training and communication
13	**Continuous improvement**	Continuous improvement	Continuous information quality improvement	Continuous improvement	Continuous improvement	Continuous improvement
14	**Information architecture management**		Information architecture management			
15	**Information security management**		Information security management			
16	**Storage management**		Storage management			
17	**Risk management**		Information quality risk management	Risk management	Risk management	
18	**Physical environment**			Physical environment	Physical environment	Conducive physical environment/continuo us power supply
19	**Understanding of the systems and DQ**			Understanding of the systems and DQ		Understanding the system and importance of DQ

Table XIV– Critical Success Factors for Data Quality Management (continued)

Nb	Critical Success Factors	CSF for TQM	Baškarada & Koronios (2014)	Xu, Koronios & Brown (2003)	Xu (2015)	Akpon-Ebiyomare et al. (2012)
20	**Personnel competency**			Personnel competency	Personnel competency	
21	**DQ policies and standards**			DQ policies and standards	DQ policies and standards	DQ policies and standards
22	**DQ controls/Input Controls**			DQ controls/ Input Controls	DQ controls/Input Controls/Internal controls	DQ controls & improvement/Input controls/Internal controls
23	**Nature of the IS**			Nature of the IS	Nature of the AIS	Nature of IS
24	**Employee relations**			Employee relations	Effective employee relations	Employee personnel relations
25	**Management of changes**			Management of changes	Management of changes	Clange management
26	**Audit and reviews**			Audit and reviews	Audit and reviews	
27	**Evaluate cost/benefit trade-offs**		Continuous improvement	Evaluate cost/benefit trade-offs	Evaluate cost/benefit trade-offs	

Note: AIS means Accounting Information System

Table XV - Identification and Explanation of CSFs for DQM

Nb	CSF	Explanation of CSF
1	Management commitment and leadership	Top management must lay a solid foundation of clear values and policies, as well as give the necessary resources (Hietschold et al., 2014). A hierarchical structure with clear descriptions of essential roles and responsibilities is required (Baškarada & Koronios, 2014).
2	Organization for quality	A good organizational structure can help in the generation of high-quality data (Xu et al., 2003). Define responsibilities for the DQ: identify the owners and custodians; appointment of data stewards and a data champion; appoint an expert or a group of people as DQ managers (Santos, 2015).
3	Training	Employee engagement and empowerment require knowledge of data quality concepts, methods, and tools (Hietschold et al., 2014). Training needs should be recognized and documented, and training workshops should be held on a regular basis. Mentoring programs should ensure on-the-job professional development in addition to formal training (Baškarada & Koronios, 2014).
4	Data quality requirements management	Identify all relevant stakeholders, collect their requirements, then model them (Baskarada & Koronios, 2014).
5	Supplier Partnership	Data supplier quality management entails establishing a successful data quality management relationship with raw data suppliers, which consists of two key components: 1. To come to an agreement on the acceptable level of raw data quality, including requirements for availability, timeliness, accuracy, and completeness; 2. To provide data suppliers with regular data quality reports and technical assistance (Xu et al., 2003).
6	Data Product Lifecycle Management	Effective data quality management requires managing data as a product as well as efficiently managing information processes (life cycles of critical information products). Identifying and documenting the data flow within the business, as well as between the organization and any external parties, is one of the aspects of this CSF (Baškarada & Koronios, 2014). Process ownership (process owners), boundaries, and steps are all well-defined (Saraph et al., 1989).

Table XV– Identification and Explanation of CSFs for DQM (continued)

Nb	CSF	Explanation of CSF
7	Data quality assessment/monitoring	Prior to attempting any DQ improvements, the existing status of DQ must be assessed, and qualitative and quantitative metrics must be developed and used (DQ-KPI) (Baškarada & Koronios, 2014). Except for accuracy, profiling tools can be used to access most of the data quality dimensions.
8	Teamwork	Improving data quality requires employee empowerment, as well as the assignment of DQ roles (see Table XII) and collaboration between business and IT personnel (Xu et al., 2003).
9	Customer focus and satisfaction	Concentrate on the wants and demands of data clients in terms of quality. Allow users to actively participate in data quality assurance and improvement (Xu et al., 2003).
10	Strategic Data Quality Planning	To achieve consistent and long-term excellence, companies must incorporate data quality into their organizational strategy. Organizations cannot set clear goals and priorities for data consumers without a strategic quality plan (Hietschold et al., 2014).
11	Continuous Data quality management improvement	Continuous DQM improvement entails developing qualitative and quantitative DQM metrics, also known as Key Performance Indicators (DQM-KPIs) and using them to track the performance of an organization's DQM activities over time (Baškarada & Koronios, 2014).
12	Culture and communication	Encouragement of a data quality improvement culture within the organization (Black & Porter, 1996). Communication is considered a two-way process, with channels for feedback available. Communication is viewed as an ongoing process, with attention given to approaches of reinforcing the concepts in the future (Porter & Parker, 1993)
13	Continuous improvement	To retain the organization's long-term competitive advantages, ongoing focus and improvement of DQ is necessary. As a result, constant and consistent improvement of machine and human data quality controls is required (Xu et al., 2003).
14	Data architecture management	The mapping of linkages between business processes, the logical data model, software, and hardware is part of this CSF. The logical business data model must be mapped into the software layer (i.e., various physical model instances), which must then be mapped into the hardware layer (Baškarada & Koronios, 2014).

Table XV– Identification and Explanation of CSFs for DQM (continued)

Nb	CSF	Explanation of CSF
15	Data security management	Access security is a critical DQ dimension, and data security management requires the implementation of effective access controls that ensure that all users are properly authenticated and authorized with the bare minimum of privileges. IS developers, for example, should not be given access to the production environment. Audit trails (logs of users' activities on the IS) must also be analyzed (for example, for exceptions) and reviewed on a regular basis (Baškarada & Koronios, 2014).
16	Storage management	Effective storage management is essential to ensure that data are accessible in the long run (Baškarada & Koronios, 2014). Implement data reuse and preservation practices (Santos, 2015).
17	Risk management	Risk management can be defined as the awareness of the implications of poor DQ and the level of commitment to reducing them (Xu et al., 2003). DQ risks to company objectives (such as financial, reputational, and regulatory risks) must be diagnosed, documented, assessed, categorized, prioritized, and mitigated/controlled. Effective DQ Risk Management should allow organizations to concentrate their DQM efforts on the most critical information products, resulting in increased DQM efficiency and effectiveness (Baškarada & Koronios, 2014).
18	Physical Environment	This entails a comfortable physical working environment, such as a modern setting with air conditioning and sufficient workspace (Xu et al., 2003).
19	Understanding of the information systems and DQ	Personnel participating in information systems must understand how information systems work (technical competence) and the importance of data quality (Xu et al., 2003).
20	Personnel competency	The ability of employees in charge of IS is especially important; for example, exceptionally talented and knowledgeable personnel in both technical and business areas are required (Xu et al., 2003).
21	DQ policies and standards	Simple, relevant, and consistent data quality policies and standards must be in place inside the organization.

Table XV– Identification and Explanation of CSFs for DQM (continued)

Nb	CSF	Explanation of CSF
21 (cont)	DQ policies and standards (cont)	It is made up of two primary parts: 1. Establishing specified and relevant data quality policies and standards; 2. Policies and standards implementation and enforcement (Xu et al., 2003).
22	DQ controls/Input Controls	For DQ improvement activities, proper DQ controls, processes, and procedures (particularly input controls should be in place) (Xu et al., 2003).
23	Nature of the IS	According to (Xu et al., 2003) this entails having appropriate systems/packages, which comprise the following elements: 1. The system is simple and intuitive to use 2. It performs as much data validation as feasible automatically (based upon business rules, etc.) 3. Documentation is enough and sufficient for individuals to follow 4. The system is easy to modify/upgrade 5. The system is well-established (stable) 6. The system is modern (adopt new technology) 7. The system is integrated and easily interpretable 8. The system has data management strategies that work, such as a centralized database and a data warehouse
24	Employee relations	Employee self-satisfaction, job security, and career advancement are all high priorities. Employees that are happy and fulfilled generate higher-quality work (Xu et al., 2003).
25	Management of changes	This is the ability and skill of the organization to manage internal and external changes. Examples of internal changes include organizational restructurings, the introduction of new technology, and staff changes. Examples of external changes include government regulations, technology, the economy, and market changes (Xu et al., 2003).
26	Audit and reviews	It is important to conduct independent internal and external audits of the systems and data quality to ensure that appropriate controls are in place. It is also important to conduct regular data quality reviews (Xu et al., 2003).
27	Evaluate cost/benefit trade-offs	It is critical to evaluate the costs of poor DQ and corresponding improvement activities, as well as any potential benefits or cost savings that may result from any process enhancements, before moving forward with any process improvements (Baškarada & Koronios, 2014).

3 PHILOSOPHICAL PERSPECTIVE AND RESEARCH DESIGN

3.1 Introduction

In general terms, research consists of imaginative and methodical labor carried out to expand the body of knowledge about a phenomenon. This includes knowledge about people, culture, and society. Research can also come up with fresh uses for the knowledge already in existence (OECD, 2015). It is an essential procedure that entails posing and trying to address worldly questions (Dane, 1990). Considering the research paradigm that we will use in this thesis, and which will be presented later in this chapter, we will consider research as "a process through which we attempt to achieve systematically and with the support of data the answer to a question, the resolution of a problem, or a greater understanding of a phenomenon" (Hevner & Chatterjee, 2010, p. 3).

Research on information systems should be both rigorous and relevant (Benbasat & Zmud, 1999; Lucas & Palma-dos-reis, 2008; Zmud, 1996). According to (Zmud, 1996, p. xxxvii) "rigorous research means that it demonstrates a soundness regarding its theoretical and conceptual development, its methodological design and execution, its interpretation of findings, and its use of these findings in extending theory or developing new theory". Relevant research implies that it demonstrates "a meaningfulness regarding its application to the significant problems and opportunities being faced by today's organizations and their members"(Zmud, 1996, p. xxxvii). The research presented in this thesis, which aims to be both rigorous and relevant, can be considered "pragmatic science" (Anderson et al. (2001) (Figure 7).

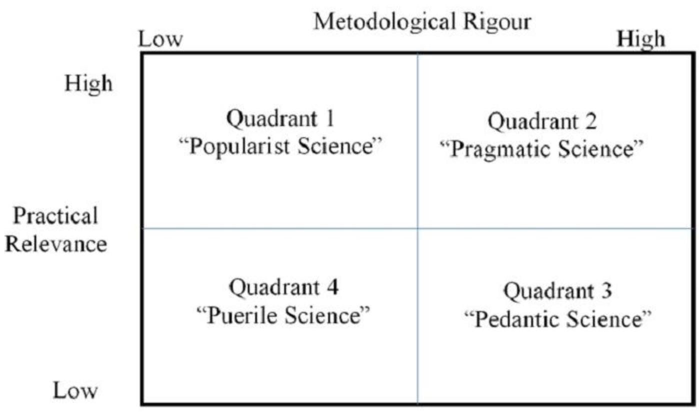

Figure 7 - Typology of Research
Source: (Anderson et al., 2001, p. 394)

3.2 Research Paradigms

Research paradigms are the ways of thinking about the world, which comprise the researcher's beliefs and philosophical assumptions about the nature of the world and how it can be understood (Alele & Malau-Aduli, 2023). The four philosophical components of research paradigms are methodology, ontology, axiology, and epistemology. These four components guide the planning and execution of research projects (Wahyuni, 2012).

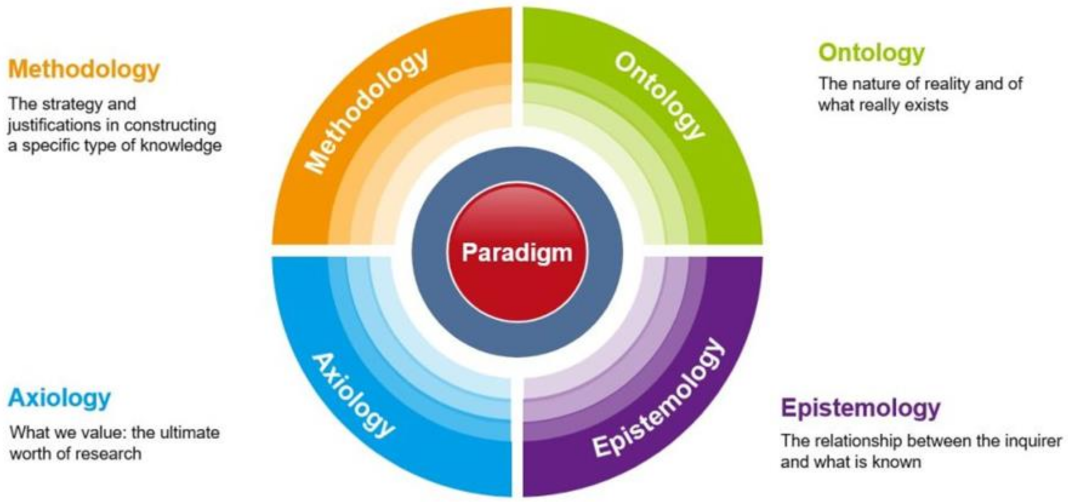

Figure 8 – Research Paradigms
Source: (Alele & Malau-Aduli, 2023)

The four components of a research paradigm are defined below.

Ontology

"Ontology is the branch of philosophy that deals with theories about the structure and behavior of the worlds that humans perceive" (Wand & Weber, 2004, p. iii). Ontology helps to conceptualize the form and nature of reality and what we believe can be known about it. It prompts us to consider things like: is the social environment a construct of one's own mind, or is there reality out there? What makes reality what it is? (Kivunja & Kuyini, 2017).

Epistemology

Epistemology encompasses assumptions regarding knowledge, what qualifies as acceptable, legitimate, and genuine knowledge, and how we might impart information to others (Saunders et al., 2019).

Axiology

The term "axiology" describes the ethical considerations that must be made when planning a research project. This entails identifying, assessing, and comprehending notions of appropriate and inappropriate behavior in relation to the research (Kivunja & Kuyini, 2017).

Methodology

The general word "methodology" refers to the research design, methods, approaches, and procedures used in a well-planned investigation to ascertain something. The methodology describes the reasoning and progression of the methodical procedures carried out in a research endeavor to learn more about a research subject (Kivunja & Kuyini, 2017).

3.3 Paradigms in Information Systems Research

Some authors refer to paradigms as philosophies, which are often used interchangeably to describe assumptions researchers make in their work (Saunders et al., 2019). The three paradigms that are most frequently applied in information systems research will be analyzed below.

Positivism

According to positivism, which is associated with the natural scientist's philosophical perspective, creating law-like generalizations requires dealing with an observable social reality, which guarantees precise and accurate knowledge. The term positivism describes the significance of what is "posited," or "given”. This highlights the positivist emphasis on a rigorously scientific empiricist approach intended to produce pure data and facts free from prejudice or human interpretation (Saunders et al., 2019).

Table XVI – The Components of Positivism

Ontology (nature of reality or being)	Epistemology (what constitutes acceptable knowledge)	Axiology (role of values)	Typical methods
	Positivism		
Real, external, independent One true reality (universalism) Granular (things) Ordered	Scientific method Observable and measurable facts Law-like generalisations Numbers Causal explanation and prediction as contribution	Value-free research Researcher is detached, neutral and independent of what is researched Researcher maintains objective stance	Typically deductive, highly structured, large samples, measurement, typically quantitative methods of analysis, but a range of data can be analysed

Source: Saunders et al. (2019, p. 144)

A positivist researcher may base his/her hypothesis development on preexisting theory. These statements provide hypothetical explanations that can be tested and confirmed, in whole or part, or refuted, which encourages the development of theory and its potential testing through additional studies. To facilitate replication, positivist researchers are likely to employ a highly structured methodology. Moreover, the focus is on measurable findings that are amenable to statistical analyses (Saunders et al., 2019).

Interpretivism

Interpretivism contends that because physical phenomena cannot be examined in the same way as people and their social environments, social science research must diverge from that of the natural sciences rather than attempt to mimic the former. Interpretivists criticize positivist attempts to find clear-cut, universal "laws" that apply to everyone because different people from different cultural backgrounds, in different situations and at different times create and experience different social realities. Rather, they think that if such complexity is completely reduced to a set of law-like generalizations, then significant insights into mankind are lost.

Table XVII – The Components of Interpretivism

Ontology (nature of reality or being)	Epistemology (what constitutes acceptable knowledge)	Axiology (role of values)	Typical methods
Interpretivism			
Complex, rich Socially constructed through culture and language Multiple meanings, interpretations, realities Flux of processes, experiences, practices	Theories and concepts too simplistic Focus on narratives, stories, perceptions and interpretations New understandings and worldviews as contribution	Value-bound research Researchers are part of what is researched, subjective Researcher interpretations key to contribution Researcher reflexive	Typically inductive. Small samples, in-depth investigations, qualitative methods of analysis, but a range of data can be interpreted

Source: Saunders et al. (2019, p. 145)

Critical Realism

Critical realists prioritize reality over all other philosophical issues, with a well-organized and multi-layered ontology being essential. Critical realists hold that although reality is external and autonomous, it is not directly accessible to us through observation or knowledge. Instead of the actual things, what we feel are some of the manifestations of the things in the real world, or what is known as "the empirical” or sensations. Thus, the goal of critical realism research is to explain observable organizational occurrences by identifying the underlying causes and mechanisms that determine the day-to-day operations of organizations through deeply ingrained social structures (Saunders et al., 2019).

Table XVIII- The Components of Critical Realism

Ontology (nature of reality or being)	Epistemology (what constitutes acceptable knowledge)	Axiology (role of values)	Typical methods
Critical realism			
Stratified/layered (the empirical, the actual and the real) External, independent Intransient Objective structures Causal mechanisms	Epistemological relativism Knowledge historically situated and transient Facts are social constructions Historical causal explanation as contribution	Value-laden research Researcher acknowledges bias by world views, cultural experience and upbringing Researcher tries to minimise bias and errors Researcher is as objective as possible	Retroductive, in-depth historically situated analysis of pre-existing structures and emerging agency Range of methods and data types to fit subject matter

Source: Saunders et al. (2019, p. 144)

Critical Realism lies between the interpretive perspective that reality is a mental construct and the positivist notion that there is a "world out there" independent of our interpretations. The goal of realistic research is to find generative mechanisms rather than predictive theories because realists argue that although the "world out there exists, it may not be feasible to perceive its essence" (Caldeira, 2000, p. 78). "Critical realists see reality as external and independent, but not directly accessible through our observation and knowledge of it" (Saunders et al., 2019, p. 147). Bhaskar (1978), cited by Caldeira (2000), distinguished three categories of reality: the empirical, the actual, and the real, to categorize experiences, events, and mechanisms (Figure 9).

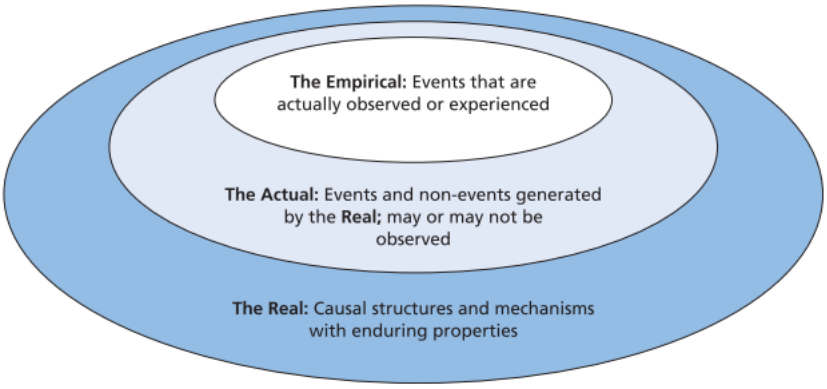

Figure 9 – Critical Realism Stratified Ontology

Source: (Saunders et al., 2019, p. 148)

In this thesis, we will apply the Critical Realism Paradigm because of our presumptions on the nature of the phenomenon being studied (a framework of CSFs for data quality management) and our perspective on the methods by which knowledge can be acquired.

3.4 Research Design

This research aims to answer two questions:

What are the critical success factors for data quality management? and

What are the priorities for implementing CSFs according to their importance?

The IS discipline sits at the intersection of people, organizations, and technology (Hevner et al., 2004). As a result, research on information systems varies regarding its purpose. The following table displays Gregor's (2006) taxonomy for the types of theory in the field of information systems research.

Table XIX - A Taxonomy of Theory Types in Information Systems Research

Theory Type	Distinguishing Attributes
I. Analysis	Says what is. The theory does not extend beyond analysis and description. No causal relationships among phenomena are specified and no predictions are made.
II. Explanation	Says what is, how, why, when, and where. The theory provides explanations but does not aim to predict with any precision. There are no testable propositions.
III. Prediction	Says what is and what will be. The theory provides predictions and has testable propositions but does not have well-developed justificatory causal explanations.
IV. Explanation and prediction (EP)	Says what is, how, why, when, where, and what will be. Provides predictions and has both testable propositions and causal explanations.
V. Design and action	Says how to do something. The theory gives explicit prescriptions (e.g., methods, techniques, principles of form and function) for constructing an artifact.

Source: (Gregor, 2006 p. 620)

To address the research questions, this research aims to create a list of CSFs for data quality management, and to determine a priority for their implementation according to their importance. These outcomes can be considered an artifact, thus setting our research in type V theory type (design and action) of table XIX. Hevner et al. (2004) characterize an artifact as a construct, a model, a method, or an instantiation. Artifacts are "broadly defined as constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), and instantiations (implemented and prototype systems)" (Hevner et al., 2004, p. 77).

Constructs "provide the vocabulary and symbols used to define problems and solutions" (Hevner et al., 2004, p. 83). As an example, the authors present the entity relationship model by Chen. Models are higher order constructions using constructs, used to describe tasks, situations, or artifacts (March & Smith, 1995) and methods are ways used to carry out goal-oriented activities (March & Smith, 1995). Instantiations might take the shape of software or intellectual tools designed to streamline the information system development process (Hevner et al., 2004).

Involving people, organizations and technology, much of the study of information systems can be divided into two categories: behavioral science and design science (Hevner et al., 2004). “The behavioral science paradigm seeks to develop and verify theories that explain or predict human or organizational behavior” (Hevner et al., 2004, p. 75), thus falling in theory types II, III, and IV of table XIX. “The design-science paradigm seeks to extend the boundaries of human and organizational capabilities by creating new and innovative artifacts” (Hevner et al., 2004, p. 75) which corresponds to theory type V of table XIX.

According to Vaus (2001) the purpose of a research design is to ensure that the evidence we collect allows us to answer the initial question as unambiguously as possible, or, put another way, a research design is a roadmap that lays out how a study will progress from its research purpose/questions to its findings (Abutabenjeh & Jaradat, 2018). Research methods refer to data collection methods (Abutabenjeh & Jaradat, 2018; Vaus, 2001) such as interviews, observation, analysis of documents, etc.

3.4.1 Design Science

Simon (1988, pp. 68–69), cited by Baskerville et al. (2015, p. 542), defines design-science as “a body of intellectually tough, analytic, partly formalizable, partly empirical, teachable doctrine about the design process”. To create knowledge, a given design problem must be analyzed, solutions must be synthesized based on this analysis, and the solution must be evaluated.

Hevner & Chatterjee (2010, p. 5) define Design Science Research as “a research paradigm in which the designer answers questions relevant to human problems via the creation of innovative artefacts, thereby contributing new knowledge to the body of scientific evidence”.

Figure 10 presents the design science framework for research in information systems.

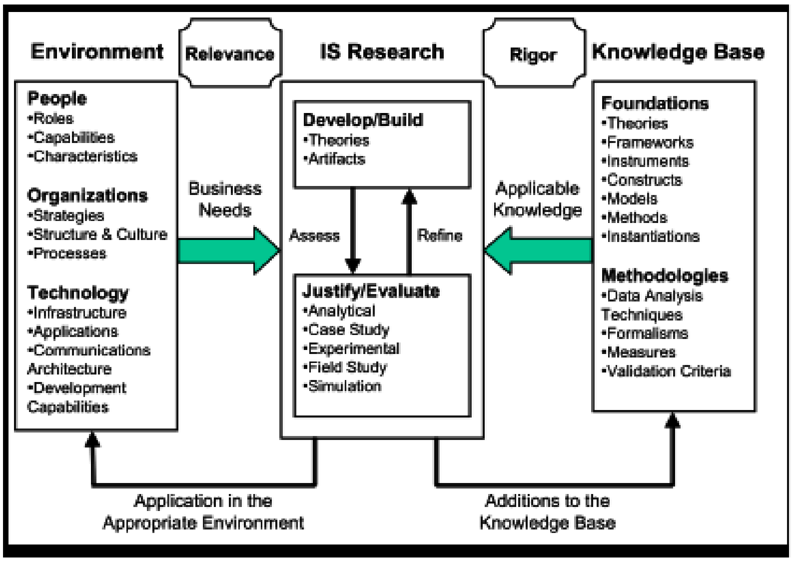

Figure 10 – Design Science Framework for Research in Information Systems

Source: (Hevner et al., 2004, p. 80)

Baskerville et al. (2015) identified four genres of inquiry in Design Science studies, which are presented in Figure 11. By "genres of inquiry" the authors mean modes of reasoning, or, put another way "style of thinking, or a manner of finding out" (Baskerville et al., 2015, p. 549). These four genres of inquiry represent all the ways of doing design-science studies, considering the two dualities of design versus science and nomothetic versus idiographic (Baskerville et al., 2015).

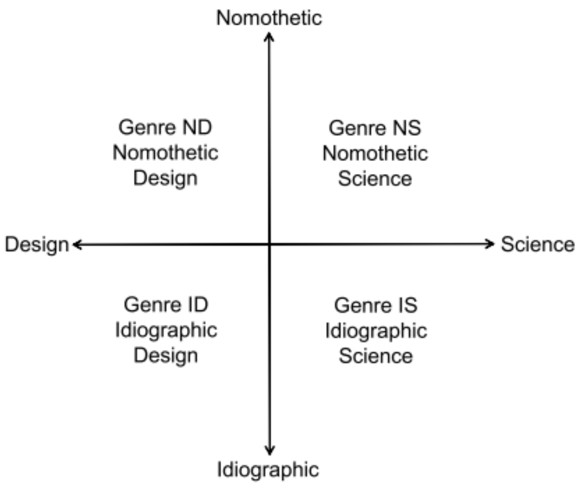

Figure 11 - Genres of Inquiry Framework for Design-Science

Source: (Baskerville et al., 2015, p. 550)

Genre of Inquiry: Nomothetic Design (ND)

Nomothetic design generates design knowledge that is relevant to a specific segment of the population. This knowledge creates a class of concrete acts that will transform a class of current circumstances into ones that are desired.

Genre of Inquiry: Idiographic Design (ID)

Idiographic design is the application of knowledge to a specific problem setting or artifact that develops a plan of action that transforms an unfavorable circumstance into a desired one.

Genre of Inquiry: Nomothetic Science (NS)

The goal of knowledge processes in nomothetic science is to generate knowledge that is both scientific and nomothetic.

Genre of Inquiry: Idiographic Science (ISc)

A systematic and validated study of a certain problem context or artifact is known as idiographic science.

The artifact created in this thesis, which consists of a list of CSFs for data quality management with a priority for their implementation according to their importance (we can call it a framework), is an artifact of type model, which according to March & Smith (1995) is a higher order construction used to describe tasks, situations, or artifacts. Given that this study is about developing knowledge applicable to “general classes of design problems” (Baskerville et al., 2015, p. 549), it should be considered in the genre of inquiry Nomothetic Science (NS).

Figure 12 presents our research design. Considering that the subject of the critical success factors for data quality management is still relatively unknown and understudied, it was decided to use a multimethod approach as proposed by (Mingers, 2001; Skulmoski & Hartman, 2007), to gradually improve the research findings.

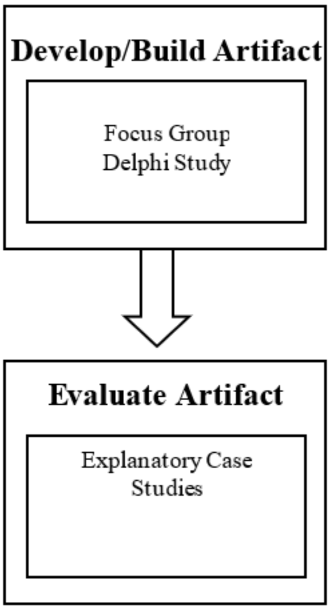

Figure 12 - Research Design

Concerning the evaluation of the artifact, Hevner et al. (2004) propose the design evaluation methods shown in Table XX below. Two explanatory case studies were selected because the aim was to study the artifact in-depth in a professional setting.

Table XX – Design Evaluation Methods

1. Observational	Case Study: Study artifact in depth in business environment
	Field Study: Monitor use of artifact in multiple projects
2. Analytical	Static Analysis: Examine structure of artifact for static qualities (e.g., complexity)
	Architecture Analysis: Study fit of artifact into technical IS architecture
	Optimization: Demonstrate inherent optimal properties of artifact or provide optimality bounds on artifact behavior
	Dynamic Analysis: Study artifact in use for dynamic qualities (e.g., performance)
3. Experimental	Controlled Experiment: Study artifact in controlled environment for qualities (e.g., usability)
	Simulation – Execute artifact with artificial data
4. Testing	Functional (Black Box) Testing: Execute artifact interfaces to discover failures and identify defects
	Structural (White Box) Testing: Perform coverage testing of some metric (e.g., execution paths) in the artifact implementation
5. Descriptive	Informed Argument: Use information from the knowledge base (e.g., relevant research) to build a convincing argument for the artifact's utility
	Scenarios: Construct detailed scenarios around the artifact to demonstrate its utility

Source: (Hevner et al., 2004, p. 86)

3.4.2 Qualitative Methods

The decision to use qualitative rather than quantitative methods stem from the importance of a deep understanding of the "socially constructed" nature of the beliefs and experiences that the people involved have about CSFs for DQM. Qualitative studies follow the process of analytical generalization (Denzin & Lincoln, 2011), and seek to understand the nature of the beliefs and experiences that the people involved have about CSFs for DQM. Furthermore, there is an intimate relationship between the researcher and what is studied (Denzin & Lincoln, 2011).

Qualitative research is "any type of research that yields results that aren't based on statistical procedures or other methods of quantification" (Strauss & Corbin, 1990, p. 17) cited by Golafshani (2003, p. 600). Instead, the research produces findings from real-world settings where the "phenomenon of interest unfolds naturally" (Patton, 2002, p. 39) cited by Golafshani (2003, p. 600). Unlike quantitative research, qualitative research may not be subject to statistical generalization, which is a method of generalizing the results to a wider, unstudied population by using a known statistical relationship between the study sample and the larger population (Yin, 2016). Qualitative studies follow the process of analytical generalization.

> *Analytic generalization may be defined as a two-step process. The first involves a conceptual claim whereby investigators show how their study's findings are likely to inform a particular set of concepts, theoretical constructs, or hypothesized sequence of events. (...). The second step involves applying the same concepts or theoretical constructs to implicate other similar situations.*
>
> (Yin, 2016, p. 105)

In the exploratory phase of the study, a Focus Group followed by a Delphi Study were used.

Focus Group

The focus group research approach was developed by Merton (1947), cited by Sobreperez (2008). A focus group is a group of individuals that meet to focus on a particular topic, conducted by a mediator (Sobreperez, 2008). The focus group research method "benefits from participant interaction, which can reveal common thoughts, reactions, and opinions on the study's topic" (Belanger, 2012, p. 110). It has been deemed useful for information systems researchers who are using exploratory research and looking for new subjects to investigate (Belanger, 2012).

Belanger (2012) prescribes that to be manageable, focus groups should be composed of 3 to 10 participants, the most common number being 5 to 7, and the meeting should last between 1 and 2 hours maximum, to avoid fatigue. In this research study, due to its exploratory nature, six participants were selected for the focus group due to their knowledge about the topic.

The Delphi Method with the Q-Sort Technique

The original Delphi method was developed by Norman Dalkey of the RAND Corporation in the 1950s for a U.S. sponsored military project. It is an iterative process that combines the opinions of a group of experts to reach a consensus.

Delphi is a method for organizing a group communication process that allows a group of people to deal with a complex problem (Linstone & Turoff, 2002). The method allows for the study of qualitative data, and it employs several experts rather than random sampling. In the context of a Delphi study, an expert is a specialist in the field of knowledge within which the study is developed. A Delphi study consists of a series of questionnaires, each one corresponding to a round. The rounds continue until a consensus is reached or it is found that one is not possible. The classical Delphi method is characterized by four key features (Rowe & Wright, 1999):

- Anonymity of Delphi participants;
- Iteration: allows the participants to adjust their opinions based on how the group's work is going from round to round;
- Controlled feedback: informs the participants of the other participant's perspectives, and provides the opportunity for Delphi participants to clarify or change their views;
- Statistical aggregation of group response: allows for quantitative analysis and data interpretation.

The first round may start with a set of open questions, or a set of questions proposed by the researcher after the literature review (Skulmoski & Hartman, 2007; Yousuf, 2007). In this specific case, the questions concerned CSFs for data quality management. In any round, experts may propose novel issues (in this case, CSFs for data quality management) relevant from their point of view.

The experts selected for this study are recognized academics and practitioners. Several agenda problems necessitated the use of Delphi online, which was supplemented by an upstream focus group and downstream case studies. Thus, a multi-method approach was used, as proposed by (Mingers, 2001; Skulmoski & Hartman, 2007).

Delphi techniques are most appropriate under two circumstances (Linstone & Turoff, 2002):

- Although the topic does not lend itself to precise analytical procedures, it can benefit from collective subjective judgments;
- Due to time or cost constraints, individuals who need to interact cannot be brought together in a face-to-face exchange (Linstone, 1981).

Given that both the above circumstances applied in this study, it was decided to use that method.

The Q-methodology was developed by Stephenson (1953), cited by Santos (2004), and it provides grounds for the systematic study of subjectivity. The distinctive feature of the Q-Sort technique, a component of Q-methodology, is that panel members must order the questions provided under a predefined distribution, usually approximately normal (Bown, 1993). The advantage of the Q-Sort scale over the Likert one is that the former reveals how participants prioritize different items (in our case CSFs) relative to each other, offering a deeper understanding of subjective preferences. Unlike a Likert scale, which rates items independently, Q-sort emphasizes the comparative rating, providing insights into the structure of participants' viewpoints (Ho, 2017).

The concept of consensus in a group is a condition of homogeneity or consistency of opinion among its members. Although consensus is a key element of Delphi studies (Schmidt, 1997), few researchers have clearly defined consensus in statistical terms, the definition being almost always arbitrary. In many studies, the decision to stop is motivated by a lack of time or by a sharp drop in the response rate between rounds, among other factors. To properly decide the round in which Delphi should halt, it is necessary to ensure the use of statistical measures to assess the consensus between members of the panel.

Given that this study asked the experts to order a list of CSFs by their relevance, Kendall's W coefficient of agreement was chosen as one of the tests of consensus determination. This coefficient determines the degree of agreement of the members of the panel. The value of this coefficient increases with the level of agreement, varying between 0 (with no consensus) and 1 (perfect consensus) (Schmidt, 1997). Schmidt provided a table to interpret different values of W, in which 0.5 indicates moderate agreement. The simple fact that members of the panel have an important level of consensus does not determine the convergence between the various rounds, so the Spearmen's rho correlation coefficient was also used between the results of two successive rounds (Santos & Amaral, 2004). These two coefficients measure not only the agreement of experts within a round but also the convergence given by the correlation between the rounds.

Case Studies

As specified above, the evaluation phase of our artifact will be implemented using explanatory case studies. The main objective of this step is to evaluate the usefulness of the developed artifact, which is a list of critical success factors (CSFs) for data quality management and a priority for their implementation, according to their importance.

According to Yin (2003, p. 13):

> *A case study is an empirical inquiry that*
>
> - *Investigates a contemporary phenomenon within its real-life context, especially when*
> - *The boundaries between phenomenon and context are not evident.*

Benbasat et al. (1987) consider that case research is particularly well suited to problems in which research and theory are still in their infancy and sticky, practice-based problems in which the actors' experiences matter and the context of action is critical. Furthermore, Yin (2003, p. 15) specifies that case study research can be used in evaluation research to “explain the presumed causal links in real-life interventions”. Similarly, Baskerville et al. (2015) consider case studies, among other methods, to be useful for artifact evaluation in the context of design science research.

4 FIELD WORK

This chapter describes the fieldwork carried out during this investigation, which consists of two phases. In the first, an exploratory phase, a focus group was used to refine the CSFs obtained through the literature review and a Delphi Study using the Q-Sort method was carried out to obtain the rating of the CSFs. The second - an explanatory phase - consisted of two case studies in companies from different industries.

4.1 Exploratory Phase

4.1.1 Focus Group

The duration of the meeting was 1 hour, and 16 minutes and the 6 attendees were recognized academics and practitioners. The role of the moderator was carried out by this author. The focus group discussion was recorded, with the participants' permission.

Prior to the meeting, the participants were sent a list of the CSFs (Table XV above) by email, and they were informed of the focus group's objective: to try to reduce the number of CSFs.

The meeting sought consensus in answer to the following three questions:

- Is there any CSF that should not be considered as such?
- Are there some CSFs missing? Which ones?
- Is there any CSF that can be merged with another or others?

The moderator began the meeting by defining the concept of CSFs for DQM as the limited number of areas in which results, if they are satisfactory, will ensure data with better quality. She then reminded the participants of the above questions to discuss.

Four participants agreed to add "data governance" to the list of CSFs, and one of them defined data governance as a "set of key actions to ensure data compliance with organizational strategies". The group also decided to include "quality organization" and "teamwork" in the CSF "data governance", thus adapting its definition.

The focus group reached a consensus that "physical environment", "storage management", "employee relations" and "nature of IS" should be considered contingency factors, not CSFs.

One participant said there were a lot of intersections in the CSF definitions and the whole group agreed with him. The moderator agreed with the group and asked the participants to help, as far as possible, to redefine the CSFs whose definitions intersect. This was partially achieved, as will be seen below.

One participant outlined the essential CSFs, from her point of view:

- Data quality should be in line with the strategic plan. This CSF was later included in the "Management Commitment and Leadership" CSF;
- DQ policies and standards;
- Monitoring and control of compliance with policies and standards;
- Input validations;
- Organizational culture with a focus on DQ;
- Top Management commitment to DQ;
- External data flows (included in CSF "Information Product Lifecycle Management"). She explained that this CSF gained great importance with the entry into force of the General Data Protection Regulation (GDPR), on May 28, 2018. This CSF has enormous relevance to respond to external reporting.

Another participant pointed out that, from his point of view, the most relevant CSFs are:

- Data Quality Assessment/Monitoring;
- Stakeholder identification (included in the CSF "Data Quality Requirements Management");
- Management Commitment and Leadership;
- Continuous Data Quality Management Improvement;
- Focus on Data Customer Satisfaction;
- Data Architecture Management, considered as the architecture of the IS ecosystem or its geography, is relevant to the type of DQ initiatives. The architecture of IS ecosystem is related to the flows of information, as well as the ownership of the data in each system.

The same participant suggested adding a CSF which he termed "company maturity concerning data processing", which was seconded by another participant. The moderator argued that "maturity" should not be considered a CSF, because it corresponds to the state of the company, before any DQ initiative, and therefore cannot be considered as "an area that must be given special and continual attention to succeed in DQ initiatives". The participant presented an example of the applicability of the CSF "company maturity concerning data processing" to a company that has just started their activity, and the moderator argued that that was a special case, and the proposed CSF should not be considered generically, so the CSF was not accepted.

Two participants suggested the CSF “strategic data quality planning” could be considered as “data quality should be in line with the strategic plan”, and one participant suggested that it could be included in the CSF “management commitment and leadership”.

It was proposed that the CSF “understanding the information systems and DQ” be renamed as “understanding the information systems and the relevance of DQ”.

The participants reached a consensus to include the CSF "continuous improvement" and “DQ controls/input controls” in the CSF "continuous data quality management improvement", defined as the “set of actions that must be taken to improve data quality”. One participant suggested including in the description of "continuous data quality management improvement” the following “set of actions we must take to improve data quality” and gave an example of a change in a data collection screen to facilitate the insertion of certain data.

After some discussion, the group agreed to include micro-changes, such as the change of an attribute domain, in the CSF "management of changes”. One participant proposed, and it was accepted, to change the definition of “management of changes” to include the following “active inclusion of the DQ requirements updating in the context of management of changes”.

Some CSFs were renamed: the CSF “understanding of the information systems and DQ” was renamed “understanding of the information systems and the relevance of DQ”, the CSF “risk management” changed to “DQ risk management”, and “customer focus and satisfaction” changed to “focus on data customer satisfaction”, considering two types of customers: clients (external customers) and the so-called users, who are internal customers. The definition of DQ-KPIs was removed from CSF “Continuous Data Quality Management Improvement” and included in “Data Quality Assessment/Monitoring”.

As a result of the meeting, the author reorganized the list of CSFs, changed some of their names, and modified some of the definitions as agreed at the focus group. The report of the focus group meeting (Appendix A), as well as the amended list of CSFs, was sent to the participants, requesting their feedback. Five of the six participants in the focus group responded to the request, of which four did not propose changes and one suggested a minor change in the report text, but not in the identification and description of CSFs. The resultant list of CSFs and their definitions are detailed in Table XXI.

Table XXI - Identification and Explanation of CSFs to DQM, Changed by Focus Group Decisions

Nb	CSF	Explanation of CSF
1	Management Commitment and Leadership	Top management must establish a solid foundation of clear data quality values and policies, as well as supply the necessary resources (Hietschold et al., 2014). "Data quality should be in line with the strategic plan" (Focus Group). To achieve consistent and long-term excellence, companies must incorporate data quality into their organizational strategy (Hietschold et al., 2014).
2	Data Governance	This entails a series of important steps to guarantee that data are compliant with organizational strategies (Focus Group). It establishes a proper organizational structure for the generation of high-quality data. It should specify who is responsible for the DQ: appoint data stewards and a data champion (Santos, 2015); appoint an expert or a group of experts as DQ managers. Promote teamwork between business and IT people, as a key to improve data quality (Xu et al., 2003) (Focus Group).
3	Training	Employee engagement and empowerment require knowledge of data quality concepts, methods, and tools (Hietschold et al., 2014). Training needs should be recognized and documented, and training workshops should be held on a regular basis. Mentoring programs should ensure on-the-job professional development in addition to formal training (Baskarada & Koronios, 2014).
4	Data Quality Requirements Management	Identify all relevant stakeholders, collect their requirements, then model them (Baskarada & Koronios, 2014).
5	Supplier Partnership	Data supplier quality management entails establishing a successful data quality management relationship with raw data suppliers. It consists of two key components: 1. To come to an agreement on the acceptable level of raw data quality, including requirements for availability, currency, accuracy, and completeness; 2. To provide data suppliers with regular data quality reports and technical assistance (Xu et al., 2003).
6	Data Product Lifecycle Management	Managing information as a product as well as effectively managing the information processes (life cycles of critical information products) is important for effective data quality management. One of the aspects of this CSF includes identifying and documenting the data flow within the organization as well as between the organization and any external parties (i.e., information product supply chain management) (Baskarada & Koronios, 2014). Clarity of process ownership (process owners), boundaries, and steps (Saraph et al., 1989).

Table XXI – Identification and Explanation of CSFs to DQM, Changed by Focus Group Decisions (continued)

Nb	CSF	Explanation of CSF
7	Data Quality Assessment/Monitoring	Prior to attempting any DQ improvements, the existing status of DQ must be assessed, and qualitative and quantitative metrics must be developed and used (DQ-KPI) (Baskarada & Koronios, 2014). Most of the data quality dimensions may be accessed using profiling tools. DQM metrics or Key Performance Indicators (DQ-KPIs) should be defined qualitatively and quantitatively, and then utilized to continually analyze the efficacy of corporate DQM activities (Baskarada & Koronios, 2014)(Focus Group). DQ should be evaluated at regular intervals using the same data profiling tools. In addition, policy and standard compliance should be monitored.
8	Focus on Data Customer Satisfaction	This means concentrating on the needs and quality requirements of data clients. It should allow data clients to actively participate in ensuring and improving data quality (Xu et al., 2003). Clients (external customers) and internal customers are both referred to as "data customers."
9	Continuous Data Quality Management Improvement	“Set of actions we must take to improve data quality” (Focus Group). There is a need for continuous and consistent data quality improvement, materialized as a set of actions that must be taken to improve data quality, such as input validations and human data quality controls (Xu et al., 2003).
10	Culture and communication	Encouragement of a data quality improvement culture within the organization (Black & Porter, 1996). Communication is considered a two-way process, with channels for feedback available. Communication is viewed as an ongoing process, with attention given to approaches of reinforcing the concepts in the future (Porter & Parker, 1993).
11	Data Architecture Management	Architecture of the IS ecosystem, or its geography, is relevant to the type of DQ initiatives. The architecture of the IS ecosystem should be described, namely the flows of information should be depicted, and the ownership of the data in each system should be identified (Focus Group).

Table XXI – Identification and Explanation of CSFs to DQM, Changed by Focus Group Decisions (continued)

Nb	CSF	Explanation of CSF
12	Data Security Management	Access security is a critical DQ dimension, and data security management necessitates the implementation of effective access controls that ensure that all users are properly authenticated and authorized with the bare minimum of privileges. IS developers, for example, should not be given access to the production environment. Audit trails (logs of users' activities on the IS) must also be analyzed (for example, for exceptions) and reviewed on a regular basis (Baskarada & Koronios, 2014).
13	DQ Risk Management	Risk management can be defined as the awareness of the implications of poor DQ and the level of commitment to reducing them (Xu et al., 2003). DQ risks to company objectives (such as financial, reputational, and regulatory risks) must be diagnosed, documented, assessed, categorized, prioritized, and mitigated/controlled. Effective DQ Risk Management should allow organizations to concentrate their DQM efforts on the most critical information products, resulting in increased DQM efficiency and effectiveness (Baskarada & Koronios, 2014).
14	Understanding of the IS and the relevance of DQ	It is important to understand how the information systems work (technical competence) and IT personnel and data customers need to understand the importance of data quality (Xu et al., 2003).
15	Personnel Competency	The ability of employees in charge of IS is especially important; for example, exceptionally talented and knowledgeable personnel in both technical and business areas are required (Xu et al., 2003).
16	DQ Policies and Standards	Simple, relevant, and consistent data quality policies and standards must be in place inside the organization. There are two primary parts: 1. Establishing specified and relevant data quality policies and standards; 2. Policies and standards implementation and enforcement (Xu et al., 2003).
17	Management of Changes	"Active inclusion of the DQ requirements updating in the context of management of changes" (Focus Group). DQ requirements, which can be internal or external, should be included and consistently updated in the process of management of changes. Internal changes include structural changes, such as organizational restructuring as well as micro changes, such as the change of an attribute domain. External changes include things such as government regulations, technology, economy, and market changes (Xu et al., 2003).
18	Audits and Reviews	Conduct independent internal and external audits of the systems and data quality to ensure that appropriate controls are in place and conduct regular data quality reviews (Xu et al., 2003).
19	Evaluate cost/benefit trade-offs	It is critical to evaluate the costs of poor DQ and corresponding improvement activities, as well as any potential benefits or cost savings that may result from any process enhancements, before moving forward with any process improvements (Baskarada & Koronios, 2014).

4.1.2 Delphi Study

The initial version of the questionnaire was submitted to two data quality experts to check its readability and to correct any inconsistencies. In addition, the Delphi online questionnaire was delivered to a panel of DQ specialists to classify the CSFs. The questionnaire included the 19 CSFs mentioned above. They were presented to the participants in alphabetical order with their associated descriptions (see Table XXI). Unfortunately, there was a space at the beginning of the CSF "Data Governance", so it was presented first, and it was the sole break to the alphabetical sequence.

The 94 DQ experts were selected from data quality professionals, acting in various industries all over the European Union (89%) and academics with work on the subject (11%). Most of the experts were found on the platform Linkedin, searching for “data quality”, “data analysis”, and “data management”. Some others were selected from the author’s acquaintances working in the field of data quality.

With Q-Sort the panel members must look at the factors as a whole and divide them into three groups: the most important; the least important and those of neutral value, according to a pre-defined normal distribution. Subsequently, they must focus on each of the lists separately and choose the most and the least important of the remaining factors, depending on the list in question. This way a list sorted by order of importance is obtained, with no ambiguity of classification and without repeated factors in the same position.

Figure 13 shows an example of the Q-Sort table for 19 questions.

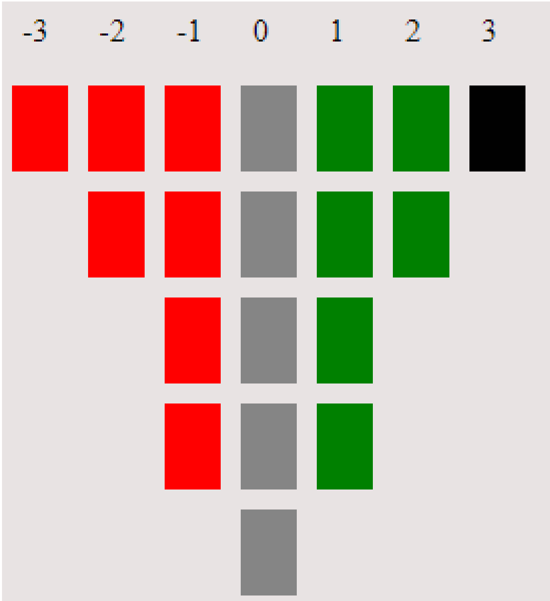

Figure 13 - Q-Sort table for 19 questions

Source: The author

First Round

For the first round, which lasted two weeks, 94 experts were asked to sort the CSFs in order of most important to least important. Fifty-six answered, which corresponds to a response rate of 60%.

As a result of the first round, the factors were ordered by increasing sum of points, with 1 point assigned to the most important factor and 19 points to the least important. In this round the experts were also asked to suggest factors that were not represented on the list submitted to them. Two new CSFs were proposed, which were not considered because they were included in CSFs presented in the initial list. The result of the first round is presented in Table XXII.

To evaluate the consensus among the participants, Kendall's W coefficient of the agreement was used. Its value was 0.195, significant at 0.000, which, according to Schmidt (1997), indicates very weak agreement. For that reason, it was decided to launch a second round.

Table XXII - Results of the First Round

Position	Sum of Points	Average	Variance	Standard Deviation	FCS Order	CSF
1	300	5,45	25,77	5,08	1	Data Governance
2	341	6,2	14,13	3,76	3	Continuous Data Quality Management Improvement
3	384	6,98	15,54	3,94	7	Data Quality Assessment/Monitoring
4	413	7,51	31,77	5,64	14	Management Commitment and Leadership
5	429	7,8	23,42	4,84	5	Data Architecture Management
6	456	8,29	24,62	4,96	8	Data Quality Policies and Standards
7	523	9,51	36,07	6,01	4	Culture and Communication
8	534	9,71	24,06	4,91	9	Data Quality Requirements Management
9	563	10,24	29,07	5,39	11	Data Security Management
10	581	10,56	22,1	4,7	16	Personnel Competency
11	588	10,69	30,48	5,52	13	Focus on Data Customer Satisfaction
12	590	10,73	23,02	4,8	10	Data Quality Risk Management
13	603	10,96	27,41	5,24	6	Data Product Lifecycle Management
14	627	11,4	32,65	5,71	2	Audit and Reviews
15	639	11,62	24,2	4,92	18	Training
16	642	11,67	20,41	4,52	15	Management of Changes
17	657	11,95	26,2	5,12	19	Understanding of the IS and the relevance of Data Quality
18	745	13,55	14,62	3,82	12	Evaluate Cost/Benefit Trade-offs
19	835	15,18	21,56	4,64	17	Supplier Partnership

Second Round

The resultant order of CSFs from the first round was sent to the respondents of the first round, and they were asked to respond to a new round, given the very weak consensus obtained. The second round lasted for three weeks, and only the 56 experts who responded to the first round were invited to participate. The results of the second round are presented in Table XXIII.

Table XXIII - Results of the Second Round

Position	Sum of Points	Average	Variance	Standard Deviation	FCS Order	CSF
1	99	3,41	16,68	4,08	1	Data Governance
2	174	6	19,93	4,46	14	Management Commitment and Leadership
3	192	6,62	14,96	3,87	3	Continuous Data Quality Management Improvement
4	204	7,03	20,46	4,52	5	Data Architecture Management
5	219	7,55	23,9	4,89	8	Data Quality Policies and Standards
6	229	7,9	28,6	5,35	4	Culture and Communication
7	243	8,38	26,89	5,19	7	Data Quality Assessment/Monitoring
8	274	9,45	30,61	5,53	9	Data Quality Requirements Management
9	298	10,28	24,78	4,98	16	Personnel Competency
10	301	10,38	23,39	4,84	13	Focus on Data Customer Satisfaction
11	310	10,69	21,58	4,65	11	Data Security Management
12	333	11,48	11,04	3,32	15	Management of Changes
13	339	11,69	34,22	5,85	2	Audit and Reviews
14	341	11,76	18,62	4,31	6	Data Product Lifecycle Management
15	341	11,76	21,69	4,66	18	Training
16	350	12,07	15,42	3,93	10	Data Quality Risk Management
17	363	12,52	24,33	4,93	19	Understanding of the IS and the relevance of Data Quality
18	414	14,28	18,99	4,36	12	Evaluate Cost/Benefit Trade-offs
19	486	16,76	8,4	2,9	17	Supplier Partnership

The second round was answered by 29 experts, which corresponds to a response rate of 52%, lower than in the first round. The Kendall's W coefficient of the agreement was 0.317, significant at 0.000, which according to Schmidt (1997) indicates weak agreement. The Spearmen's rank correlation coefficient between the results of the first and the second rounds (Santos & Amaral, 2004) was 0.674, significant at 0.01 level, which indicates a good positive correlation. The Spearman's rho correlation coefficient ranges between [-1,1], with -1 representing perfect negative monotonic correlation and +1 representing perfect positive monotonic correlation, 0 means no correlation.

The CSF Data Governance held first place in both rounds and the last three CSFs also maintained the same positions in the two rounds. This result reflects the theory associated with the Q-Sort which states that the panel members are more confident on the most and least critical issues (Santos & Amaral, 2004).

The results in Table XXIV show that the first 11 CSFs of the 1st round remain in the first 11 places in the 2nd round, although, except for Data Governance ranked first, they have changed their positions.

Table XXIV - Results of the First and Second Rounds

CSF Numbers Ordered by First Round	CSF Descriptions ordered by First Round	CSF Numbers Ordered by Second Round	CSF Descriptions ordered by Second Round	Order in the Second Round
1	Data Governance	1	Data Governance	1
3	Continuous Data Quality Management Improvement	14	Management Commitment and Leadership	2
7	Data Quality Assessment/Monitoring	3	Continuous Data Quality Management Improvement	3
14	Management Commitment and Leadership	5	Data Architecture Management	4
5	Data Architecture Management	8	Data Quality Policies and Standards	5
8	Data Quality Policies and Standards	4	Culture and Communication	6
4	Culture and Communication	7	Data Quality Assessment/Monitoring	7
9	Data Quality Requirements Management	9	Data Quality Requirements Management	8
11	Data Security Management	16	Personnel Competency	9
16	Personnel Competency	13	Focus on Data Customer Satisfaction	10
13	Focus on Data Customer Satisfaction	11	Data Security Management	11
10	Data Quality Risk Management	15	Management of Changes	12
6	Data Product Lifecycle Management	2	Audit and Reviews	13
2	Audit and Reviews	6	Data Product Lifecycle Management	14
18	Training	18	Training	15
15	Management of Changes	10	Data Quality Risk Management	16
19	Understanding of the IS and the relevance of Data Quality	19	Understanding of the IS and the relevance of Data Quality	17
12	Evaluate Cost/Benefit Trade-offs	12	Evaluate Cost/Benefit Trade-offs	18
17	Supplier Partnership	17	Supplier Partnership	19

Considering the weak consensus obtained and despite fearing a sharp drop in the number of responses, it was decided to hold the third round.

Third Round

The CSF ordering in the 2nd round was sent to the respondents, and they were asked to respond to a new round, given the still weak consensus obtained. The 3rd round lasted for three weeks, and only the 29 experts who responded to the 2nd round were invited to participate.

The results of the third round are presented in Table XXV.

The third round was answered by 18 experts, which corresponds to a response rate of 62%, higher than in the first and second rounds. The higher response rate was the result of insistence with the experts.

Table XXV - Results of the Third Round

Position	Sum of Points	Average	Variance	Standard Deviation	CSF Order	CSF
1	55	3,44	18,13	4,26	1	Data Governance
2	74	4,63	9,18	3,03	14	Management Commitment and Leadership
3	79	4,94	8,06	2,84	3	Continuous Data Quality Management Improvement
4	109	6,81	17,63	4,2	5	Data Architecture Management
5	111	6,94	13,53	3,68	4	Culture and Communication
6	112	7	14,27	3,78	8	Data Quality Policies and Standards
7	126	7,88	23,05	4,8	7	Data Quality Assessment/Monitoring
8	155	9,69	11,56	3,4	9	Data Quality Requirements Management
9	159	9,94	26,46	5,14	13	Focus on Data Customer Satisfaction
10	164	10,25	26,07	5,11	6	Data Product Lifecycle Management
11	179	11,19	28,16	5,31	16	Personnel Competency
12	179	11,19	15,9	3,99	15	Management of Changes
13	190	11,88	18,52	4,3	11	Data Security Management
14	203	12,69	10,36	3,22	10	Data Quality Risk Management
15	206	12,88	42,65	6,53	2	Audit and Reviews
16	209	13,06	13,66	3,7	18	Training
17	214	13,38	25,85	5,08	19	Understanding of the IS and the relevance of Data Quality
18	237	14,81	10,83	3,29	12	Evaluate Cost/Benefit Trade-offs
19	279	17,44	5,46	2,34	17	Supplier Partnership

The Kendall's W coefficient of the agreement was 0.433, significant at 0.000, indicating weak to moderate agreement (Schmidt (1997). The Spearmen's rank correlation coefficient between the results of the second and the third rounds (Santos & Amaral, 2004) was 0.256, significant at 0.01 level, indicating weak correlation. It was intriguing that this coefficient dropped from 0.674 (good positive correlation) between the first and the second rounds to 0.256 (weak positive correlation), but we do not really know why.

The CSF Data Governance held first place in all rounds and the last three CSF (Understanding of the IS and the Relevance of Data Quality, Evaluating Cost/Benefit Trade-offs, and Supplier Partnership) also retained the same last positions in the three rounds. It should be noted that the first eight CSFs remained in the top eight positions in the three rounds, although in some cases they changed their order. Table XXVI presents the results of the three rounds.

Table XXVI – Results of the Three Rounds

CSF Numbers Ordered by First Round	CSF Numbers Ordered by Second Round	CSF Numbers Ordered by Third Round
1	1	1
3	14	14
7	3	3
14	5	5
5	8	4
8	4	8
4	7	7
9	9	9
11	16	13
16	13	6
13	11	16
10	15	15
6	2	11
2	6	10
18	18	2
15	10	18
19	19	19
12	12	12
17	17	17

CSF Numbers and Description	
1	Data Governance
2	Audit and Reviews
3	Continuous Data Quality Management Improvement
4	Culture and Communication
5	Data Architecture Management
6	Data Product Lifecycle Management
7	Data Quality Assessment/Monitoring
8	Data Quality Policies and Standards
9	Data Quality Requirements Management
10	Data Quality Risk Management
11	Data Security Management
12	Evaluate Cost/Benefit Trade-offs
13	Focus on Data Customer Satisfaction
14	Management Commitment and Leadership
15	Management of Changes
16	Personnel Competency
17	Supplier Partnership
18	Training
19	Understanding of the IS and the relevance of Data Quality

To better understand the results, it was decided to perform a cluster analysis of the variables "sum of points" and "standard deviation" using Ward's method. The dendrogram obtained is shown in Figure 14, where the numbers of the CSFs shown in Figure 14 correspond to their order in Round 3. The dendrogram presents three clusters with good intra-cluster distance, the first one consisting of CSFs 1-7, the second 8-13 and the third 14-19.

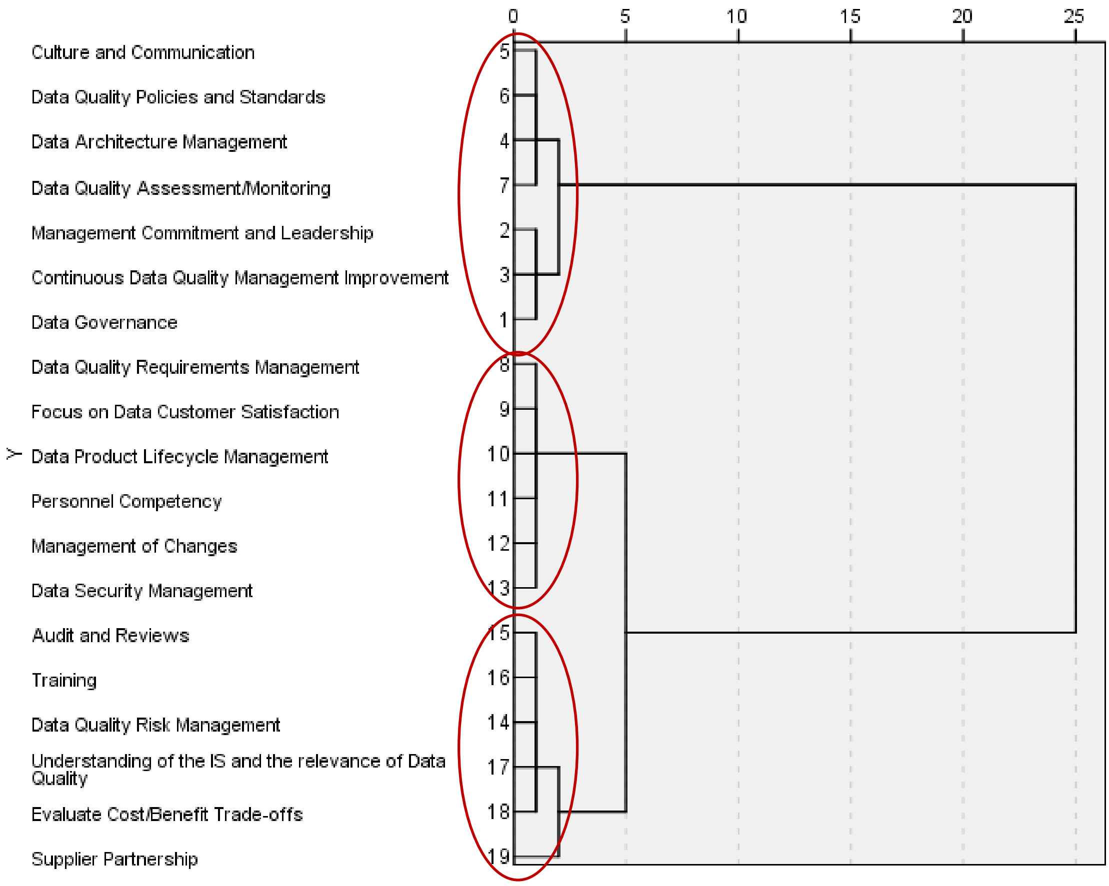

Figure 14 - Dendrogram obtained using Ward's Method for Hierarchical Clusters

The first 10 CSFs that were found, except for Data Governance, are in line with results from previous research (Akpon-Ebiyomare et al., 2012; Baškarada & Koronios, 2014; Santos, 2015; Williams et al., 2015; Xu, 2015; Xu et al., 2003). The consistent and clear rating of Data Governance in the first place is noteworthy, for previously it was only identified in three papers (Baskarada & Koronios, 2014; Santos, 2015; Santos & Lucas, 2019). Baskarada & Koronios (2014, p. 285), citing Bell et al. (2008, p. 4), identified "Information Quality Management Governance", defined as "the specification of decision rights and an accountability framework to encourage desirable behavior in the valuation, creation, storage, use, archival and deletion of information". This definition corresponds to that of data governance. According to Khatri & Brown (2010, p. 149) the term "data governance refers to who holds the decision rights and is held accountable for an organization's decision-making about its data assets" while management "involves making and implementing decisions" (Khatri & Brown, 2010, p. 148).

Thus, “Information Quality Management Governance” is a misconception because governance and management mean slightly different things, whether in terms of data or any other asset.

The first position of data governance is probably due to the increasing importance of data to organizations, particularly in the context of Regulatory Compliance, Advanced Analytics and Data Science, which have given rise to the implementation of a Data Governance approach and the function of Chief Data Officer among other new roles (Table XXII above). It should also be noted that the first three CSFs - Data Governance, Management Commitment and Leadership and Continuous Data Quality Management Improvement - clearly stand out from the rest (see Table XV).

4.2 Explanatory Phase – Case Studies

4.2.1 Methodology

The case study companies were selected from the researcher's network, and they were assured that the names of their company and interviewees would remain confidential, facilitating their acceptance of the invitation. The two companies - OwnFly and MyBank, both fictitious names - are both medium-sized enterprises located in Southern European Latin Countries. OwnFly operates as an airline while MyBank functions as a bank. In addition, the companies had to meet two more criteria:

- To have data quality initiatives in place;
- To be of different industries, allowing for theoretical replication, because we expect “contrasting results but for predictable reasons” (Yin, 2003, p. 47). Xu & Lu (2003) report that companies in different industries do not place equal importance on some CSFs.

The researcher conducted case studies, some on site, some online, with each of the participants involved, and data were collected through semi-structured interviews with key DQM players, both from IT and the business. Individual interviews with each participant lasted no more than an hour, because they were very busy people, and their roles are detailed in each case study.

The principal areas and questions covered in the interviews were the following:

- Introduction to the study and data collection;
- Anonymity of the study;
- When did the data quality initiatives start and what were the initial drivers? What are the current business drivers for data quality improvement?

- What are the current data quality improvement initiatives?
- Which is your company organization for data quality?
- What are the CSFs for DQM from your perspective?
- Are there any ongoing big data projects? What about data quality initiatives in those projects?

All the interviews were recorded, with the permission of the interviewees, and subsequently transcribed. The transcriptions were sent by email to the participants, asking them to change them if any information was incorrect. In addition to the interviews, interviewees were also contacted by the researcher via email to gather additional information, and the websites of the companies were consulted for more information.

The quality of the case studies research design was confirmed by the four tests, and their tactics, prescribed by Yin (2003): construct validity, internal validity, external validity, and reliability. The definitions of the tests and their tactics were obtained from Yin (2003, pp. 34–39).

- Construct validity – determining appropriate operational measures for the concepts under study. Tactic: various sources of evidence were used, and interviewees were invited to review their transcriptions. Partially achieved, because the drafts of the case studies were not previously sent to the interviewees;
- Internal validity – proving a causal relationship - that is, demonstrating how one set of conditions causes another, as opposed to forming fictitious associations. Tactic: matching the findings of the case studies with the results of the Delphi study and the literature review;
- External validity – “establishing the domain to which a study's findings can be generalized” (Yin, 2003, p. 34) to some broader theory. Tactic: the use of two case studies with the same research objective;
- Reliability - proving that a research's processes, including the data gathering methods, can be repeated in the same case study and yield the same findings. Tactic: the procedures used in the case studies were documented and a case study database was created using MaxQda 2022. However, it is likely that in a repetition of the case studies, the same results would not be obtained because organizations are continually improving their data quality management.

4.2.2 Data Analysis

According to Yin (2016) data analysis is carried out in five analytic phases: 1. Compiling, 2. Disassembling, 3. Reassembling, 4. Interpreting, 5. Concluding (Figure 15). To support the work, the MaxQda 2022 Computer Assisted Qualitative Data AnalysiS (CAQDAS) software was used.

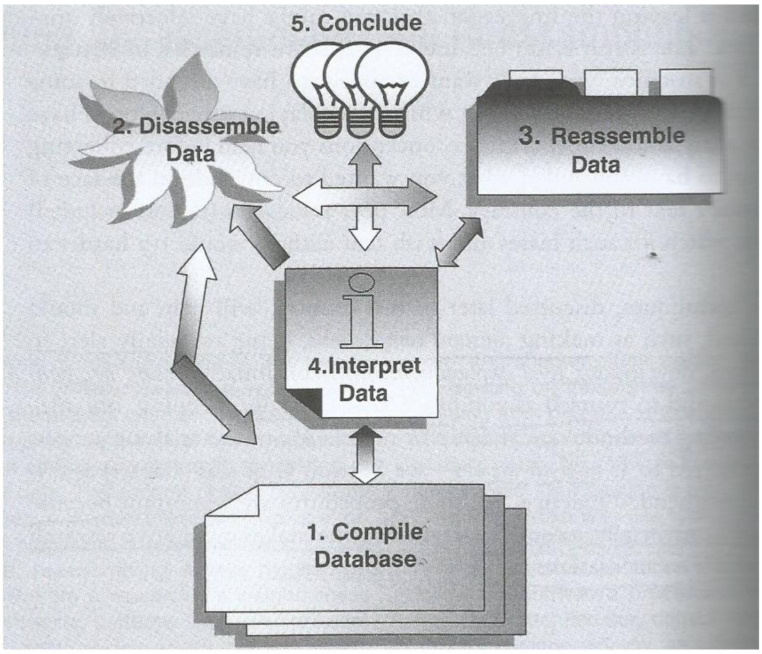

Figure 15 - Five Phases of Data Analysis
Source: (Yin, 2016, p. 186)

4.2.3 Compiling and Disassembling

The compiling phase consisted of re-reading and analyzing the interviews to understand whether new critical success factors appeared (what really happened) and importing them into the CAQDAS MaxQda 2022 (Yin, 2016).

In the disassembling phase, a hybrid approach of inductive and deductive coding (Fereday & Muir-Cochrane, 2006) was used to code the interviews with the 19 critical success factors (CSFs) previously identified and 6 new codes that emerged from the interviews (Fereday & Muir-Ccochrane, 2006). The Code Book is presented in Appendix C. The identification and definitions of new codes that emerged from the interviews are presented in Table XXVII below. The new code (CSF) sponsorship is, in fact, referred to in DAMA (2017, p. 559) as "Executive Sponsorship" for Data Management Organization and not specifically for Data Quality Management. However, the reference was not included in the literature review, where only works on CSFs for data quality management or total quality management were selected, the DAMA book being much more wide-ranging.

Table XXVII - New Codes that Emerged from the Interviews

Nb	CSF	Explanation of CSF
20	Business Processes Change	Flexibility to change business processes, using data quality processes close to the customer (Ownfly data stewards)
21	Information Catalog	Create an information catalog that supports efficient metadata management and must contain Technical definitions and Business Concepts to be queried by some users. For Business Concepts it can be a dictionary of terms (organizacional concepts and their synonyms) (OwnFly and MyBank)
22	DQ Tools	Data quality tools are the processes and technologies for identifying, understanding and correcting flaws in data that support effective information governance across operational business processes and decision making. The packaged tools available include a range of critical functions, such as profiling, parsing, standardization, cleansing, matching, enrichment and monitoring (Gartner, n.d.) (OwnFly and MyBank)
23	Sponsorship	Someone at C-Level that has a vision of the importance of information assets quality (OwnFly and MyBank)
24	Management of User Expectations	Avoid false expectations (Ownfly CIO)
25	Business Model Definition	A well-defined business model allows us to identify what needs to be improved in data quality (Ownfly CIO)

4.2.3.1 MyBank (MB)

MyBank (MB) is a fictitious name for a medium size bank located in a Southern European Latin Country in the European Union, with approximately five million clients, around 6,500 employees, and total assets between €50 billion and €100 billion. It is active in a variety of sectors, including commercial banking, investment banking, venture capital, asset management, specialized credit, and real estate. MB has mature customer data quality management, and they do not have big data projects, which will only happen with the new data architecture they are currently working on. We interviewed the Chief Information Officer (CIO), the Chief Data Officer (CDO), the person responsible for data management and quality (RDQ), who reports to the CDO, and three professionals of the Operations Center (OC), including the Chief Operations Center (COC). The Operations Center is currently the owner of the Customer Database.

At MyBank, on both the IT and business sides, there is great concern about improving the quality of customer data, as well as some pride in the progress that has already been made. The reason for this success is undoubtedly due to the management commitment and leadership and the extensive training given to branch employees throughout the country. No other data bases were identified with data quality improvement initiatives, so DQM is only developed at the business-unit level, which, according to the TDWI (Kobielus, 2022), indicates that the company is at the intermediate level of DQM maturity.

Data quality initiatives began in the 1990s with the normalization of customer names and addresses to avoid returning mail. Despite the bank's digital transformation, they currently spend around five million euros a year on mail. Employees across the country were trained in the importance of good quality data. Nowadays they invest in customer data quality in all fields required by the Regulators (National Central Bank and European Central Bank).

According to the CIO, "Data Quality Management (DQM) at MyBank comprises a set of DQ initiatives". For example, "A Chief Data Officer has been appointed, two forums have been created, a more operational data forum and a government forum", and the data quality project on the customer database. Notwithstanding these initiatives, MyBank does not have a systematic strategy for communicating the improvements they are making to the quality of customer data to their users.

The CDO is a young engineer who seemed enthusiastic about data and anything that could improve their quality. He was appointed in the first semester of 2019, and reports to the chief executive officer (CEO). According to the CDO, "My area drives change: it creates vision and strategy; it creates frameworks that support day-to-day operations that will make it possible to introduce areas of improvement at the data level".

There are four areas reporting to the CDO, namely Strategy and Data Transformation, Data Management and Quality, Data Services, and a Central Pool of Data Scientists, which will be presented using his words.

- Strategy and Data Transformation

  MyBank has accumulated a significant amount of data over the past 20-30 years, having a structured data warehouse area that no longer adequately meets current needs.

  It has DQ governance and management problems, very disparate technology that is not appropriate to the current reality, for example many COBOL applications, and data extraction systems from the mainframe that are no longer agile. MyBank needs a vision of data innovation to support the ongoing digital transformation program: an integrated vision of who our customers are, working analytics, things *that the current architecture does not support. It is important to have a structured vision to define a modern reference architecture, processes that allow us to govern data, catalog data, quality-controlled processes with KPIs, following the data lifecycle. Look at the data strategically because it is a sector heavily regulated, and data are one of our most important assets.*

- Data Management and Quality

This area will bring the know-how from the central to the distributed systems in the new architecture. [There is an] opportunity for MyBank to have governed processes, data catalog, businesses owners identified, unified concepts, a single cross-cutting view of multiple domains and data elements. It is not yet known who will own the customer data domain that is currently owned by the Operations Center.

- Data Services

Technicians implementing Extract, Transform and Load (ETL) processes, data development and operation.

- Central Pool of Data Scientists

The business has and will always have analysts. Data scientists (DSi) have knowledge of data engineering and can interpret business needs by applying a set of tools. As business areas are not mature enough to realize what a DSi is, MyBank created a centralized pool whose elements can be channeled to the business, which is where they should be. DSis are technology agnostics and currently use tools like R, Pyton and Azure (as an integrative services platform). MyBank DSis are young people and like to experiment with new technology.

The Chief Operations Center (COC), who is the owner of the customer database, stressed the importance of training the agencies' staff:

We trained employees on the importance of data and provided training across the country. The trainees were the collaborators of the network, and the trainers were the people who knew about the Customer database. They often trained trainers. The agencies selected an employee to train, who would then train the remaining colleagues.

With the intensive training program for the network, we noticed a significant decrease in the creation of duplicate customers, and, at the same time, a considerable improvement in the quality and accuracy of the data entered in the system.

Currently, the Business Drivers to data quality management, according to the interviewees, are: understanding the importance to the business of the data at its disposal (COC, CIO); improving customer relationship (CIO); regulation - ensuring consistency of various reports, namely to achieve compatibility with BCBS239 (Principles for Effective Risk Data Aggregation and Risk Reporting, 2013b) (CIO, RDQ), and Cost Optimization (CIO).

The interviewees referred to Critical Success Factors in the following words: “People at different levels, capable and involved” (CIO); “dedicated, trained, aware people with lean communication in the process” (CDO); “teams involved and aware” (COC); “End to End Commitment: From Operational to Informational people” (CDO); “Data Quality Processes must be well defined and embedded in normal operation” (CIO); “Sponsorship” (COC, CDO) – “Have an insight of the importance of the quality of information assets from the highest level” (CDO); Organization – “Model that governs data from top to bottom, that allows for certified and endowed quality data”; “Consolidation of concepts, unification of metrics, metadata management, ability to oversee DQ with people assigned” (CDO). Data Quality Management and Governance: “We need to define a set of rules, procedures, corporate practices, vision of what quality processes are used, how we monitor, how we track systems, how we keep [the] data lifecycle intact with a vision of extraction of the information value. Certified and quality data. Set of practices that must be implemented at various levels” (CDO).

4.2.3.2 OwnFly (OF)

OwnFly is a fictitious name for a medium size airline located in a Southern European Latin Country in the European Union. It is a Star Alliance member[6], has over 7,500 employees, operates one of the youngest fleets in the world, and in 2023 flew more than 17 million customers across its network and transported around 100,000 tons of cargo. OwnFly adopts a

[6] The Star Alliance is the world's largest global airline alliance.

collaborative approach to tackling sustainability challenges, aiming to maintain a transparent positive relationship with its stakeholders.

The company has started its data quality management program (as they call it) with the implementation of its second Customer Master Data Management (MDM). Indeed, according to DAMA (2017), master data are, by definition, among the most important data in any organization, and they are frequently where data quality improvement begins. Their first MDM (MDM0) was developed in-house with multiple limitations. For example, the matching process[7] worked in batch. "The MDM0 experience helped a lot to define the requirements for the MDM tool", in the words of the Data Stewards, and it was a "Proof of Concept" for the MDM Project Leader.

We interviewed the Chief Information Officer (CIO), the MDM Project Leader (PL), the Data Stewards (DS), the Chief of Customer Service (CCS), who is the Owner and Sponsor of Customer Master Data, and the IT Business Partner for Marketing and Sales (BP).

The IT Business Partner for Marketing and Sales (BP) identified her role as:

> *At OwnFly, within IT, we are organized as follows: IT Director, more internal areas, competence centers (matrix) and a more business-oriented area, and I am a business partner (small IT director) for the area of marketing and sales*

Both the CIO and the MDM Project Leader consider that data quality management is a program at the company, not a project with a beginning and an end. This aligns with the recommendations outlined in the TDWI reports (Kobielus, 2022, 2024), which advise treating Data Quality Management (DQM) as a continuous journey. Kobielus (2024, p. 19) states that "a highly mature and successful DQM practice is more of a moving target than an end state".

We noted a significant dedication among all the interviewees to improving the quality of customer data. However, there appears to be resistance towards enhancing the quality of data gathered by the sales staff. This may be because no training was given to these professionals regarding the importance of customer data quality for the company, and the company is only now considering that training. Furthermore, approximately 50% of their sales originate from

[7] Matching process is the procedure of matching the new data (for instance new reservations) with the customers that already exist in the MDM.

the Travel Agency channel, which collects data within the Amadeus system, where the fields for collecting customer data are unstructured, and there is a lack of standardization in gathering data in a machine-readable format.

As with MyBank, no other data bases were identified with data quality improvement initiatives, so DQM is developed at the business-unit level, which, according to the TDWI (Kobielus, 2022), indicates that the company is at the intermediate level of DQM maturity.

The data stewardship area depends on the architecture and looks at the customer's DQ problems. In the new MDM the sources go through a Data Quality (DQ) tool from Informatica, do the lineage[8], and update the MDM in near real time.

The Chief of Customer Service (CCS) and the CIO pointed out that airlines do not own their distribution channel, considering that their business is identical to that of large distribution. For the CCS:

> *We have a difficult problem to solve: an important source of data is from Amadeus and the way contacts are placed in Amadeus are unstructured. For example, the Travel Agent instead of putting a +34 before the telephone number, puts ES, and they can put the ES at the beginning or at the end, I don't even know if they do this so that the number is not automatically processed*

OwnFly absolutely needs to identify and contact its customers to notify them in case of irregularities, as well as to sell more. It is particularly important for the customer to feel that he/she is recognized, for example, by acknowledging that OwnFly knows that the person had an irregularity and is asking for a certain thing, such as an upgrade.

For data stewards:

> *Our Reservation system is from Amadeus, our Contact Center is from Salesforce, and it is difficult, if not impossible, to change their frontend, so validations are carried out downstream.*

About implementing an Information Catalog, the CIO considers that:

---

[8] Data lineage is the process of tracking the flow of data over time to provide a clear understanding of where the data originated, how it has changed, and its ultimate destination within the data pipeline (IBM, n.d.).

It is mandatory to correctly define business concepts (e.g. customer, revenue - various definitions, irregularity")

and Data Stewards:

What we are doing now is mapping business terms in the different systems: our dictionary of terms (concepts and their synonyms) has approximately six hundred entries and will be able to be queried by some users

The catalog is not supported by any commercial software despite having made several attempts to buy one. In the CIO's words, "What was good was extremely expensive and what was cheap was worthless".

Before reaching OwnFly, the MDM Project Leader (PL) worked at a Telecommunications Company that had a so-called canonical model, so when an attribute was needed in more than one system, it had to belong to the canonical model: business definition and technical definition. He says that at OwnFly even the technical definition is not standardized, and he believes that metadata should be born in the operational area.

For the MDM Project Leader (PL):

It is quite easy for users to say, "I need the data to be of good quality", but it is very difficult to change some point, from input, e.g., for the data to reach this quality. There is more investment in bad data to make it good than to stop bad data from entering the system. There is a cultural issue in changing business processes because of DQ

The Chief of Customer Service (CCS) considers it difficult to raise awareness in the business area for Data Quality (DQ). He said:

We have a website that sells many million euros a year and the pressure for them to sell more is remarkably high, so DQ is not in the first line of concern. We already have some Key Performance Indicators (KPI) that focus on completeness: the percentage that have email, phone, etc. With MDM we must define process KPIs: for instance, what percentage is going to the data stewards (DS)

The Data Stewards (DS) have five years of experience with customer data and are extremely focused on analyzing the data lifecycle and lineage and figuring out where anomalies could have happened.

The OwnFly communication strategy focuses on a weekly meeting between the data stewards and the business: Digital, Loyalty and Customer Service.

The following Critical Success Factors emerged for the interviewees: Continuous Data Quality Management Improvement – Input Validations (CIO, PL, CCS); Data Stewardship (BP, DS, PL); Data Product Lifecycle Management (DS); Focus on data customer satisfaction (CIO); Communication -The frontline matters (CIO, CCS); Data Quality Policies & Standards (CIO, PL, DS); Management of Changes (CIO); Business Model Definition (CIO); Data Quality Assessment/Monitoring (CIO, CCS); Personnel Competency (BP); Sponsorship (BP, PL, DS); Information Catalog (business definition and technical definition) (CIO, PL, DS); Communication (DS, PL); Supplier Partnership (CCS); Management of User Expectations (CIO, CCS); Training (CCS); Data Security Management (CCS), Audits and Reviews (CCS). For the CIO, Data Governance is a concept that includes several components, one of which is Data Quality, as stated by (DAMA, 2017; Khatri & Brown, 2010) (see Figures 5 and 6 above).

4.2.4 Reassembling and Interpreting

The reassembling process reassembles the shards or pieces into new groups and sequences what was in the original notes by using substantive themes, based on combinations of fragmented components (Yin, 2016). The art of interpreting might be defined as providing the researcher's own interpretation of the data and data arrays that have been reconstructed (Yin, 2016).

To start, we generated concept matrices using MaxQda 2022. The initial matrix, shown in Table XXVIII below, illustrates the application of codes (in this instance, corresponding to Critical Success Factors) across the interviews conducted with the representatives from MyBank and OwnFly. It should be noted that the following codes have been broken down, because the interviewees expressly mentioned them:

- Continuous Data Quality Management Improvement → Input Data Validations;
- Data Governance → Data Stewardship;
- Data Architecture Management → Data Ownership.

Table XXVIII below presents the use of codes by the two organizations with the CSFs ranked in descending order of the number of mentions. The disaggregated CSFs are highlighted in yellow and the new CSFs in green.

The variability in code usage between the two companies suggests that both the Critical Success Factors (CSFs) and their respective importance should be evaluated within the specific industry context where data quality management is implemented. As Xu & Lu (2003) demonstrated regarding Accounting Information Systems, companies in different industries do not place equal importance on some CSFs.

The first rank, with 18 mentions, includes the CSFs Continuous Data Quality Management Improvement and Information Catalog. While the former is in line with Delphi results, the latter was not included in the Delphi study. Nor is it included in the literature analyzed, although Baskarada (2009) includes the Metadata Management indicator at level 3 - "Measuring" of the DQM Capability Maturity Model. The scale has five levels, which increase as the DQM becomes more mature. Level 3 indicates that "metadata is managed separately from regular information" (Baskarada, 2009, p. 150). For the DQM Maturity Levels see Figure 16 below. Khatri & Brown (2010) consider metadata one of the five decision domains for data governance (Figure 6) and DAMA (2017) similarly considers it one of the components of Data Governance (Figure 5). Nevertheless, although the CSF Information Catalog is not presented in the research literature, it appears as a "Key Success Factor" in a professional report, TDWI (Transforming Data with Intelligence) (Kobielus, 2022), which was based on an extensive survey to more than 1,000 professionals.

Of the CSFs ranked in the first third of the table, with between 18 and 10 mentions, apart from the new codes, only DQ Risk Management was not ranked in the first third of the last round of Delphi, occupying a position further down.

Looking at the second third of the table for CSFs with between 9 and 4 mentions and removing the new code (Sponsorship), four of the CSFs were not ranked in the second third of the last round of Delphi: Evaluate cost/benefit trade-offs (ranked lower), Data Governance (significantly higher), DQ Policies and Standards (higher), and Management Commitment and Leadership (significantly higher).

Considering the third part of the table and removing the new CSFs, only Data Quality Requirements Management and Management of Changes were ranked higher in the Delphi.

Table XXVIII - Use of Codes in the Two Cases

Codes	OwnFly	MyBank	Total	Rank
Information Catalog	11	7	18	1
Continuous Data Quality Management Improvement (CDQMI)	13	5	18	1
CDQMI --> Input Data Validations	6	5	11	3
DQ Tools	8	3	11	3
DQ Risk Management	3	8	11	3
Culture and communication	9	2	11	3
DG -->Data Stewardship	7	3	10	7
DAM-->Data Ownership	5	5	10	7
Data Architecture Management (DAM)	1	9	10	7
Sponsorship	6	3	9	10
Focus on Data Customer Satisfaction	7	2	9	10
Personnel Competency	2	6	8	12
Data Quality Assessment/Monitoring	5	3	8	12
Data Product Lifecycle Management	2	6	8	12
Evaluate cost/benefit trade-offs	0	7	7	15
Data Governance (DG)	2	5	7	15
DQ Policies and Standards	3	3	6	17
Management Commitment and Leadership	0	4	4	18
Training	1	2	3	19
Supplier Partnership	1	2	3	19
Data Quality Requirements Management	3	0	3	19
Management of Changes	1	1	2	22
Business Processes Change	2	0	2	22
Business Model Definition	2	0	2	22
Understanding of the IS and the relevance of DQ	1	0	1	25
Management of User Expectations	1	0	1	25
Data Security Management	1	0	1	25
Audits and Reviews	1	0	1	25

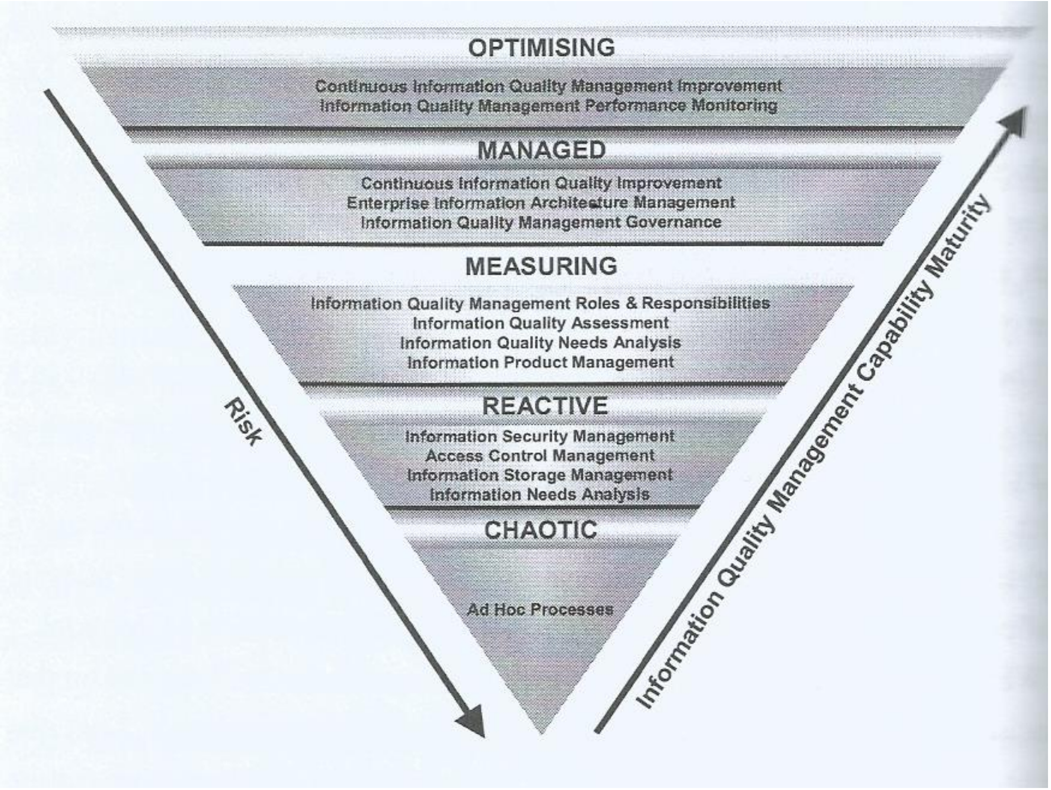

Figure 16 – DQM Capability Maturity Model

Source: Baskarada (2009, p. 226)

It is worth highlighting that the CSFs Information Catalog and DQ Tools were very well placed, in ranks 1and 3 respectively while the ranking of the CSFs Data Governance and Management Commitment and Leadership seemingly declined abruptly.

The fall in rank of the CSF Management Commitment and Leadership is perhaps not surprising given that five C-level people were interviewed: three at MyBank (CIO, CDO and COC) and CIO and CCS at OwnFly, and all of them are completely committed to data quality. For them Management Commitment and Leadership would be a given, with no need to mention it.

The lower position of the CSF data governance could be related, among other things, to the following factors:

- In Delphi, the CSF Data Governance included Data Stewardship, which was highly ranked in the case studies, but disaggregated in their analysis;
- Data Quality is theoretically considered a subset of Data Governance (as mentioned by OwnFly's CIO). See Figures 6 and 7;

- The companies that participated in the case studies did not have a Data Governance program in place.

Tables XXIX and XXX present the relative importance of CSFs in both companies. The results show that the relative importance of the CSFs differs between the two companies.

The results for OwnFly (Table XXIX) show that the most important critical success factors are, in descending order, Continuous Data Quality Management Improvement, Information Catalog, Culture and Communication, DQ Tools, Data Stewardship, Focus on Data Customer Satisfaction, Input Data Validations, Sponsorship, Data Ownership and Data Quality Assessment/Monitoring.

For MyBank (Table XXX), the most important critical success factors are, in descending order, Data Architecture Management, DQ Risk Management, Information Catalog, Evaluate Cost/Benefit trade-offs, Personnel Competency, Data Product Lifecycle Management, Continuous Data Quality Management Improvement, Input Data Validations, Data Ownership, Data Governance and Management Commitment and Leadership.

Some of the differences can be explained by the fact that the two companies operate in hugely different industries, corroborating Rockart's (1979) findings that CSFs can differ between companies, even in the same industry, and between managers, although he suggests that a few CSFs are determined by industry-specific characteristics, whereas the remaining ones are produced by variations in environmental conditions, temporal considerations, geographic location, or strategic circumstances.

It is worth noting that the two top-ranking CSFs in MyBank – Data Architecture Management and DQ Risk Management – are ranked far lower in the CSFs for OwnFly. The importance placed on Data Architecture Management may be due to the fact that MyBank has a Chief Data Officer, which is not the case at OwnFly. Regarding the CSF ranked second, DQ Risk Management may be perceived as very important in MyBank because the company operates in a highly regulated sector.

Table XXIX - Relative Importance of CSFs for OwnFly

Codes	OwnFly
Continuous Data Quality Management Improvement (CDQMI)	13
Information Catalog	11
Culture and communication	9
DQ Tools	8
DG -->Data Stewardship	7
Focus on Data Customer Satisfaction	7
CDQMI --> Input Data Validations	6
Sponsorship	6
DAM-->Data Ownership	5
Data Quality Assessment/Monitoring	5
DQ Risk Management	3
DQ Policies and Standards	3
Data Quality Requirements Management	3
Personnel Competency	2
Data Product Lifecycle Management	2
Data Governance (DG)	2
Business Processes Change	2
Business Model Definition	2
Data Architecture Management (DAM)	1
Training	1
Supplier Partnership	1
Management of Changes	1
Understanding of the IS and the relevance of DQ	1
Management of User Expectations	1
Data Security Management	1
Audits and Reviews	1
Evaluate cost/benefit trade-offs	0
Management Commitment and Leadership	0

Table XXX - Relative Importance of CSFs for MyBank

Codes	MyBank
Data Architecture Management (DAM)	9
DQ Risk Management	8
Information Catalog	7
Evaluate cost/benefit trade-offs	7
Personnel Competency	6
Data Product Lifecycle Management	6
Continuous Data Quality Management Improvement (CDQMI)	5
CDQMI --> Input Data Validations	5
DAM-->Data Ownership	5
Data Governance (DG)	5
Management Commitment and Leadership	4
DQ Tools	3
DG -->Data Stewardship	3
Sponsorship	3
Data Quality Assessment/Monitoring	3
DQ Policies and Standards	3
Culture and communication	2
Focus on Data Customer Satisfaction	2
Training	2
Supplier Partnership	2
Management of Changes	1
Data Quality Requirements Management	0
Business Processes Change	0
Business Model Definition	0
Understanding of the IS and the relevance of DQ	0
Management of User Expectations	0
Data Security Management	0
Audits and Reviews	0

While both companies recognized the importance of most CSFs, there were some exceptions. The following CSFs were mentioned by only one of the organizations:

- OwnFly: Data Quality Requirements Management, Business Processes Change, Business Model Definition, Understanding of the IS and the relevance of DQ, Management of User Expectations, Data Security Management and Audits and Reviews. It should be noted that the CSFs Business Processes Change, Business Model Definition and Management of User Expectations are not presented in the literature and were introduced by OwnFly interviewees.
- MyBank: Evaluate Cost/Benefit Trade-Offs and Management Commitment and Leadership, with the absence of the latter explained above.

To obtain a visual representation of the relative importance of the various CSFs, a Cloud Code was drawn up for both organizations (Figure 17), MyBank (Figure 18) and OwnFly (Figure 19).

Figure 17 - Cloud Code for Both Organizations

Figure 18 - Cloud Code for MyBank

Figure 19 – Cloud Code for OwnFly

To find patterns, a matrix was created to show the coding of the interviews with the CIOs of both organizations (Table XXXI). A similar matrix was drawn up for the customer database owners (Table XXXII). In both cases, codes not used in the interviews by any of the stakeholders were removed.

Table XXXI - Comparing the CIOs Interviews

Codes	OwnFly	MyBank	Total
Information Catalog	4	2	6
Data Architecture Management	0	4	4
Input Data Validations	2	1	3
Culture and communication	2	1	3
DQ Policies and Standards	1	2	3
Data Governance	2	0	2
Data Quality Assessment/Monitoring	2	0	2
Focus on Data Customer Satisfaction	1	1	2
Continuous Data Quality Management Improvement	1	1	2
Data Ownership	1	1	2
DQ Risk Management	0	2	2
Management of Changes	1	1	2
Evaluate cost/benefit trade-offs	0	2	2
DQ Tools	1	1	2
Business Model Definition	2	0	2
Data Stewardship	0	1	1
Data Product Lifecycle Management	0	1	1
Personnel Competency	0	1	1
Management of User Expectations	1	0	1

A comparison of the results shows that Information Catalog is the critical success factor (CSF) most mentioned by OwnFly's CIO, and it is also mentioned by MyBank's CIO.

By contrast, CSF Data Architecture Management is brought up several times by MyBank's CIO, but not by OwnFly's CIO, the former stating, "To be compliant with the various regulations (GDPR, BCBS239, others) we must have the data architecture defined". This position suggests that this option is due to the bank's obligation to comply with regulations. The importance of complying with regulations may also explain the fact that CSF DQ Risk Management was mentioned only by MyBank's CIO.

Both CIOs mentioned the following CSFs: Input Data Validations, Culture and Communication, DQ Policies and Standards, Focus on Data Customer Satisfaction, Continuous Data Quality Management Improvement, Data Ownership, Management of Changes and DQ Tools. Finally, although CSF Data Stewardship was not mentioned in the interview with OwnFly's CIO, the company has implemented a data stewardship function.

Table XXXII below compares the interviews with the Customer Database Owners from each organization.

Table XXXII - Comparing Customer Database Owners' Interviews

Codes	OwnFly	MyBank	Total
Continuous Data Quality Management Improvement	6	2	8
Input Data Validations	3	3	6
DQ Risk Management	2	3	5
Evaluate cost/benefit trade-offs	0	5	5
Focus on Data Customer Satisfaction	3	1	4
Data Stewardship	2	1	3
Supplier Partnership	1	2	3
Data Quality Assessment/Monitoring	2	1	3
Data Ownership	1	2	3
Training	1	1	2
Culture and communication	2	0	2
Personnel Competency	0	2	2
Information Catalog	2	0	2
Sponsorship	0	2	2
Data Governance	0	1	1
Data Quality Requirements Management	1	0	1
Data Product Lifecycle Management	0	1	1
Data Security Management	1	0	1
Audits and Reviews	1	0	1
DQ Tools	1	0	1

The results show that there is significant variation in the importance attributed to CSFs by customer database owners, with just over half the CSFs being mentioned by both organizations, and of these, only three CSFs received multiple mentions by both customer database owners. The CSF Continuous Data Quality Management Improvement received the most mentions by OwnFly, and it is also mentioned more than once by MyBank. Input Data Validations and DQ

Risk Management are similarly well considered by both owners. The CSFs Focus on Data Customer Satisfaction, Data Stewardship, Supplier Partnership, Data Quality Assessment/Monitoring, Data Ownership and Training are mentioned by both owners.

Of the CSFs that were mentioned by only one organization, it should be noted that Evaluate Cost/Benefit trade-offs received multiple mentions by MyBank's Customer Owner, who said "We take up the quality of all data that allow MyBank to get in touch with the customer: telephone, mobile, email. We invest in all the mechanisms that allow us to keep that data up to date". This same CSF was also mentioned by the CIO of MyBank.

To understand the relations among the CSFs that emerged from the case studies, a cluster analysis was carried out. Figure 20 shows the code relations for the full results from the case studies.

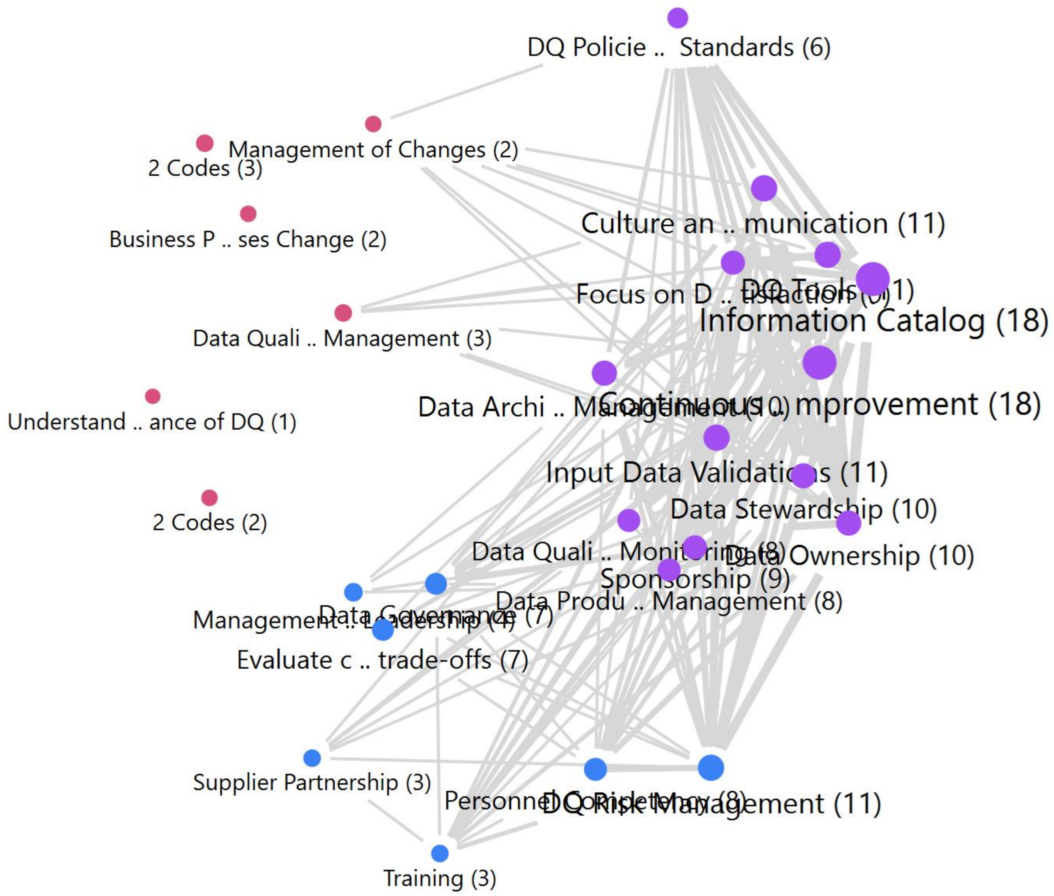

Figure 20 – Code Relations with Subcode Breakdown

An analysis of the elements of Figure 20 and Table XXXIII below shows that there are three clusters of CSFs, which we will call A, B and C. Cluster A is made up of CSFs that were very often mentioned in the interviews and that are highly related to each other. Cluster B is made up of CSFs that were somewhat mentioned in the interviews, but less than those in Cluster A, relatively related to each other, and very closely related to Cluster A. Cluster C is made up of unrelated CSFs, which were rarely used in the interviews, some of which relate to some of the CSFs in Cluster A. The outlier CSFs Audits and Reviews, Data Security Management, Management of User Expectations and Understanding of the IS and the relevance of DQ, in addition to only being used once in the interviews, have virtually no relationships with other CSFs.

Notwithstanding this analysis, we moved the CSF Management Commitment and Leadership from Cluster C to Cluster A, since we have witnessed effective commitment of management, in both companies and at various levels, in efforts to improve data quality. In addition, the CSF was rated second in the Delphi study.

Table XXXIII – Code Clusters with Breakdown of Subcodes

Codes	Cluster	Nb of Coded Segments
Management Commitment and Leadership	A	Meaningless (4)
Continuous Data Quality Management Improvement (CDQMI)	A	18
Information Catalog	A	18
CDQMI --> Input Data Validations	A	11
DQ Tools	A	11
Culture and communication	A	11
DG -->Data Stewardship	A	10
DAM-->Data Ownership	A	10
Data Architecture Management (DAM)	A	10
Sponsorship	A	9
Focus on Data Customer Satisfaction	A	9
Data Quality Assessment/Monitoring	A	8
Data Product Lifecycle Management	A	8
DQ Policies and Standards	A	6
DQ Risk Management	B	11
Personnel Competency	B	8
Evaluate cost/benefit trade-offs	B	7
Data Governance (DG)	B	7
Training	B	3
Supplier Partnership	B	3
Data Quality Requirements Management	C	3
Management of Changes	C	2
Business Processes Change	C	2
Business Model Definition	C	2
Understanding of the IS and the relevance of DQ	C	1
Management of User Expectations	C	1
Data Security Management	C	1
Audits and Reviews	C	1

To be able to compare the results of the case studies with those of the Delphi study, it was decided to aggregate the subcodes into their respective codes, the result of which is shown in

Table XXXIV and Figure 21 below. In Appendix B code relations are shown, the number in the intersection between two codes represents the number of documents (interviews) in which they were used simultaneously.

An analysis of Table XXXIV shows that there are still three clusters of CSFs, which we will call A, B and C. The clusters differ somewhat from the previous scenario, with Cluster A having nine CSFs, B six and C ten. As in the previous scenario, the CSFs in cluster A were very often mentioned in the interviews and are highly related to each other. Cluster B CSFs were somewhat mentioned in the interviews, but less than those in Cluster A, relatively related to each other, and very closely related to CSFs of Cluster A. CSFs in Cluster C were rarely used in the interviews, and some of them relate to some of the CSFs in Cluster A. The CSFs Audits and Reviews, Data Security Management, Management of User Expectations and Understanding of the IS and the relevance of DQ, in addition to only being used once in the interviews, have virtually no relationships with other CSFs. Like the procedure for finalizing the clusters of the codes with the subcodes broken down, the CSF Management Commitment and Leadership was promoted from Cluster C to Cluster A since we have witnessed the effective commitment of management, in both companies and at various levels, in efforts to improve data quality. Moreover, the CSF was rated second in the Delphi study. The results indicate that there is greater compatibility between the ranking associated with the case studies and the rating obtained in the Delphi study. The increased compatibility is to be expected because we are now comparing the same concepts.

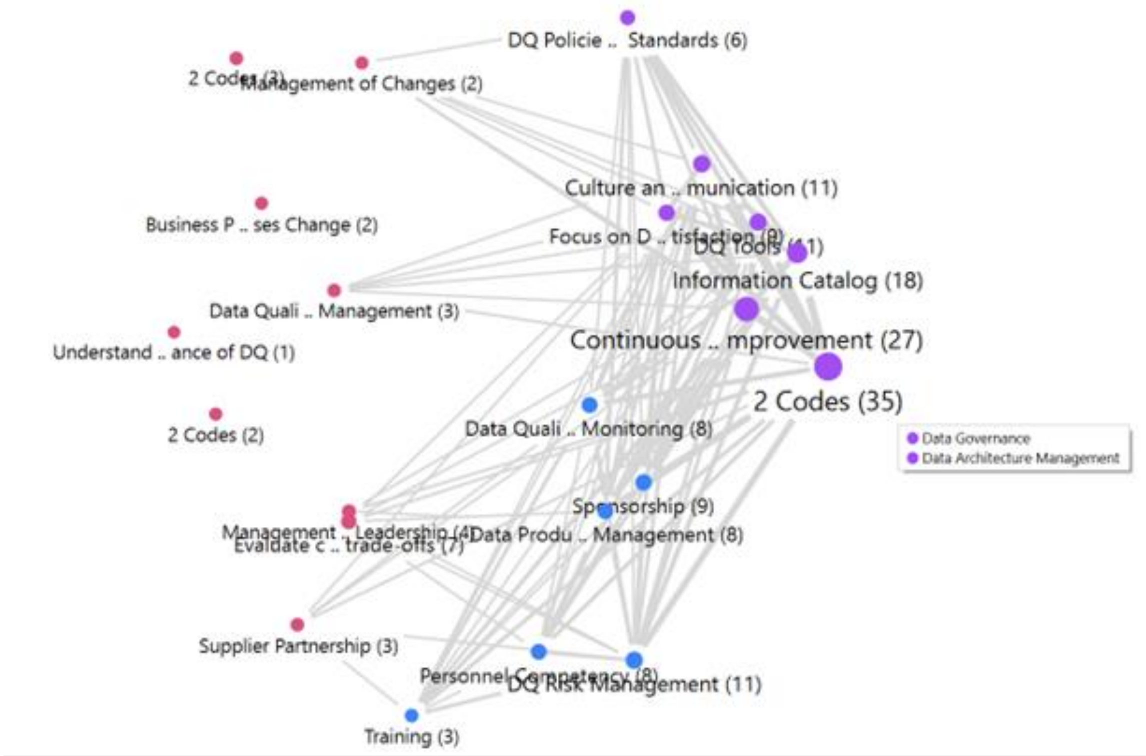

Figure 21 – Code Relations Without Breakdown of Subcodes

Table XXXIV – Code Clusters without Breakdown of Subcodes

Codes	Cluster	Position in Delphi	Nb of Coded Segments	Rank
Management Commitment and Leadership	A	2	Meaningless (4)	1
Continuous Data Quality Management Improvement (CDQMI)		3	27	2
Information Catalog			18	3
Data Architecture Management (DAM)		4	18	3
Data Governance (DG)		1	17	5
DQ Tools			11	6
Culture and communication		5	11	6
Focus on Data Customer Satisfaction		9	9	8
DQ Policies and Standards		6	6	9
DQ Risk Management	B	14	11	1
Sponsorship			9	2
Data Quality Assessment/Monitoring		7	8	3
Data Product Lifecycle Management		10	8	3
Personnel Competency		11	8	3
Training		16	3	6
Evaluate cost/benefit trade-offs	C	18	7	1
Supplier Partnership		19	3	2
Data Quality Requirements Management		8	3	2
Management of Changes		12	2	4
Business Processes Change			2	4
Business Model Definition			2	4
Understanding of the IS and the relevance of DQ		17	1	7
Management of User Expectations			1	7
Data Security Management		13	1	7
Audits and Reviews		15	1	7

The CSFs highlighted in gray bear little or no importance in the data quality management process, which was confirmed in the Delphi study and the case studies. Thus, they can be ignored. The CSFs highlighted in yellow belong to the first cluster of Ward's method (Figure 14 above), in green to the second one, and in blue to the third. Comparing the clusters relating to the case studies and Delphi, only cluster A practically coincides with the 1st Delphi cluster, which is not the case with clusters B and C.

In the interviews conducted as part of the case studies, the question "What are the current business drivers to data quality improvement?" was asked. As a result, a by-product of this stage was the identification of those business drivers, which are shown in the following table.

Table XXXV – Business Drivers for DQM

Business Driver	MyBank	OwnFly	Total
Regulations	11	1	12
Customer Satisfaction	3	6	9
Increase in Sales	4	2	6
Cost Optimization	4	0	4
Managing the Business	1	0	1
Digital Transformation	1	0	1

The business drivers identified are in line with the literature (DAMA, 2017; Kobielus, 2022; Redman, 1996), including the four most important ones, which are Regulations, Customer Satisfaction, Increase in Sales, and Cost Optimization.

4.2.5 Concluding

This study found two new CSFs, Information Catalog and DQ Tools, that were not included in the research literature and are of fundamental importance for the DQM process, as analyzed above. Although these two CSFs are not presented in research literature, they appear in a professional report, TDWI (Transforming Data with Intelligence) (Kobielus, 2022) , which was based on an extensive survey of more than 1,000 professionals. The report states that those who said they had managed data quality successfully were more likely to say they planned to use a data catalog for DQM soon. A more recent TDWI report (Kobielus, 2024) pointed out the need to use DQ tools with embedded machine learning models.

In addition, three clusters of CSFs, A, B and C, were identified in descending order of importance. Thus, the results suggest that the CSFs in each cluster should be applied in ascending order of the rank (Table XXXIV and Figure 21 above).

This study also revealed that both the CSFs for Data Quality Management and their relative importance depend on each organization and probably the industry, corroborating results from (Xu & Lu, (2003). Given such variation, the identification and prioritization of CSFs for DQM in different industries constitutes a very interesting topic for future research.

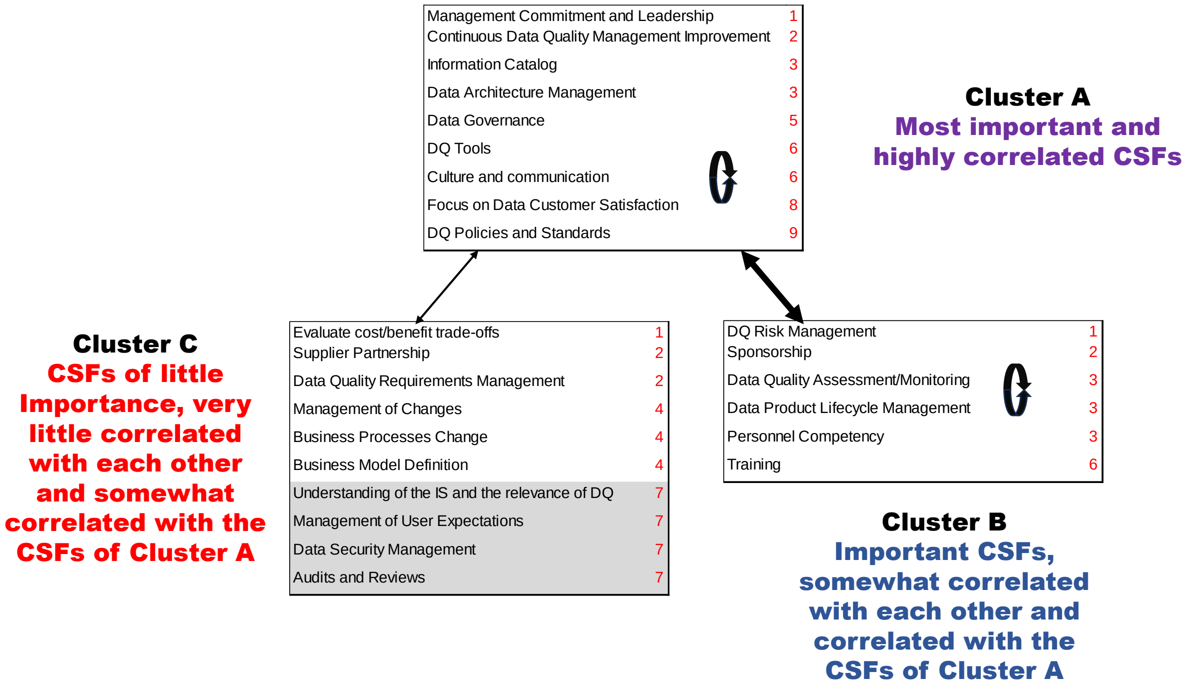

Figure 22 - CSF Clusters

Source: The author

5 DISCUSSION

The case studies only partially confirmed the order of importance of the CSFs found in the Delphi study, but they did make it possible to identify two new CSFs to which the interviewees gave great importance overall, namely Information Catalog and Data Quality Tools. It should be noted that these two CSFs are not referenced in the research literature.

Looking at the Table XXXIV above, we can see:

- The 9 CSFs in Cluster A were in the top 9 in the Delphi study (except for Information Catalog and DQ Tools, which were not analyzed in Delphi), so the results were largely confirmed for this cluster;
- For Cluster B, the results partially confirmed those of Delphi, with the exception of the CSFs "Data Quality Assessment/Monitoring" and “Training”, which were ranked respectively higher and lower in Delphi;
- For Cluster C, the case studies also partially confirmed the Delphi results: the CSF "Data Quality Requirements Management" and "Management of Changes" were ranked respectively in the 8th and 12th positions in the Delphi, far higher than the positions they are now in. The CSFs "Data Security Management", "Understanding of the IS and the relevance of DQ" and "Audits and Reviews" were also near the bottom of the Delphi.

In short, from this study it can be concluded that the nine CSFs in cluster A are the most reliable for use by organizations.

The "Data Governance" CSF deserves some reflection as it has only been identified in the literature in three documents (Baskarada & Koronios, 2014; Santos, 2015; Santos & Lucas, 2019), the first calling it “Information Quality Management Governance”. This denomination is unsuitable because management and governance are two different concepts. According to Khatri & Brown (2010, p. 148) governance refers to “what decisions must be made to ensure effective management and use” of an asset “and who makes the decisions” and management “involves making and implementing decisions”.

While the CSF Data Governance was introduced by 4 of 6 attendees of the focus group and came first in all three Delphi rounds, its definition, which was used in the Delphi and the case studies, suggests that it may be a core capability and not a critical success factor for data quality management. Indeed, DAMA (2017, p. 493) argues that “A data quality program is more effective when part of a data governance program. Often data quality issues are the reason for establishing enterprise-wide data governance”. Thus, it may be considered a core capability for data quality management (DAMA, 2017; Karkošková, 2023). This subject cannot be clarified in this work, but only in further studies.

6 RELEVANCE, CONCLUSIONS, LIMITATIONS AND FURTHER RESEARCH

6.1 Relevance

Redman (2017) states that increased data quality leads to reduced errors, reduced expenses, improved decisions, and superior outcomes for organizations. He estimates that the cost of bad data to be 15% to 25% of revenue for most companies. In 2016, IBM reported that the annual cost of bad data quality in the United States was $3.1 trillion (Redman, 2016) and Gartner (Sakpal, 2021) estimated in 2021 that, on average, poor data quality costs organizations $12.9 million per year.

The impact of poor data quality is terrible for business and certainly forces companies to improve the quality of their data. According to Tom Peters (DAMA International, 2009, p. 1) "Organizations that do not understand the overwhelming importance of managing data and information as tangible assets in the new economy will not survive".

The growing use of data analytics, especially for predictive analysis, as well as generative artificial intelligence tools, is leading companies to increasingly seek to improve the quality of their data and establish good DQM practices (Côrte-Real et al., 2020; Davenport & Tiwari, 2024; Kobielus, 2024; Liu et al., 2021).

The identification of critical success factors for data quality management, defined as the limited number of areas in which results, if they are satisfactory, will ensure data with better quality, is thus important for companies that want, or are obliged, to improve the quality of their data. This is the main reason for this research.

6.2 Conclusions

This research was motivated by a real problem of data quality in organizations, and some gaps in the theory. It aims to provide organizations with a framework of critical success factors made up of the CSFs themselves and a priority for their implementation according to their importance. It also contributes to theory by proposing three clusters of CSFs and prioritizing their implementation, as well as identifying two new CSFs: Information Catalog and DQ Tools. Information Catalog was defined by interviewees as "a catalog that supports efficient metadata management and must contain technical definitions and business concepts to be queried by some users. For business concepts it can be a dictionary of terms (organizational concepts and their synonyms". DQ Tools is defined by Gartner (n.d.) as "processes and technologies for identifying, understanding, and correcting flaws in data that support effective information governance across operational business processes and decision making. The packaged tools available include a range of critical functions, such as profiling, parsing, standardization, cleansing, matching, enrichment, and monitoring". More recently, corporations are using AI-driven DQM automation, which according to TDWI (Kobielus, 2022, p. 13), refers to "the use of embedded machine learning models to handle one or more DQM functions like reliably, repeatedly, and accurately without needing direct human oversight and assistance".

Given the objective of the study, it was decided to use the Design Science Framework, defined as "a research paradigm in which the designer answers questions relevant to human problems via the creation of innovative artefacts, thereby contributing new knowledge to the body of scientific evidence" (Hevner & Chatterjee, 2010, p. 5). The Design Science Framework consists of two phases, the first of which involves developing/creating the artifacts/theories and the second evaluating them.

The research questions for this study were:

- What are the critical success factors (CSF) for data quality management?
- What are the priorities for the implementation of CSFs according to their importance?

Both questions were answered in this research, and Clusters A, B and C were found, whose CSFs should be used in ascending order of the cluster name and, in each cluster, in ascending order of the rank number (see Figure 22 above).

6.3 Limitations

For this work qualitative research was chosen because we wanted to study in depth the perspective of the various stakeholders on what they consider to be the critical success factors for data quality management and why. As a qualitative research study, its results can only be generalized to the theory (analytical generalization) and not to the population, which is the case with quantitative research (Yin, 2016). Finally, it was not possible to determine if "Data Governance" is a CSF for data quality management or a core capability for DQM because the companies in which the case studies were carried out had not implemented a data governance approach.

6.4 Further Research

The qualitative approach of this study is very appropriate for providing fresh perspectives for potential future research studies. The research process has brought to light some ideas about which aspects need to be investigated further.

It would be interesting to apply the CSFs pretended in cluster A in the prescribed order in some organization and during a certain time and analyze the impacts of improving data quality.

Future research could focus on the impact of recent tools using machine learning and artificial intelligence on improving the quality of company data.

It would also be interesting to explore the role of Data Governance in improving data quality management, to understand if it is a critical success factor or a core capability.

Another topic to explore is the confirmation of Data Catalog and Data Quality Tools as CSFs for data quality management, as well as their prioritization.

Another idea for future research is the update of Xu & Lu's (2003) work by investigating the identification of CSFs for DQM, as well as their prioritization, in different industries, a topic that we identified in the results of the two case studies. Finally, we suggest that future research can explore the identification and prioritization of CSFs for different groups of stakeholders (e.g. data consumers, data entry staff, data stewards, data owners).

REFERENCES

Abutabenjeh, S., & Jaradat, R. (2018). Clarification of research design , research methods , and research methodology : A guide for public administration researchers and practitioners. *Teaching Public Administration*, 1–22. https://doi.org/10.1177/0144739418775787

Addagada, T. (2024, July 25). *Data Quality: The Secret Sauce for AI and Generative AI Success*. Https://Www.Linkedin.Com/Pulse/Data-Quality-Secret-Sauce-Ai-Generative-Success-Nicola-Askham-Oozff/.

Akpon-Ebiyomare, D., Chiemeke, S. C., & Egbokhare, F. A. (2012). A Study of the Critical Success Factors Influencing Data Quality in Nigerian Higher Institutions. *African Journal of Computing & ICT*, *5*(2), 45–50.

Alele, F., & Malau-Aduli, B. (2023). *An Introduction to Research Methods for Undergraduate Health Profession Students*. James Cook University. https://jcu.pressbooks.pub/intro-res-methods-health/

Anderson, N., Herriot, P., & Hodgkinson, G. P. (2001). The practitioner- researcher divide in Industrial, Work and Organizational (IWO) psychology: Where are we now and where do we go from here? *Journal of Occupational and Organizational Psychology*, *74*(391–411).

Baskarada, S. (2009). *Information Quality Management Capability Maturity Model*. GWV Fachverlage GmbH.

Baskarada, S., & Koronios, A. (2014). A Critical Success Factor Framework for Information Quality Management. *Information Systems Management*, *31*(4), 276–295. https://doi.org/10.1080/10580530.2014.958023

Baskerville, R. L., Kaul, M., & Storey, V. C. (2015). Genres of Inquiry in Design -Science Research : Justification and Evaluation of Knowledge Production. *MIS Quarterly*, *39*(3), 541–564.

Batini, C., Rula, A., Scannapieco, M., & Viscusi, G. (2015). From Data Quality To Big Data Quality. *Journal of Database Management*, *26*(1), 60–82.

Batini, C., & Scannapieco, M. (2006). Data Quality - Concepts, Methodologies and Techniques. In *Springer*. Springer-Verlag BerlinHeidelberg. https://doi.org/10.1007/3-540-33173-5

Belanger, F. (2012). Theorizing in Information Systems Research Using Focus Groups. *Australasian Journal of Information Systems*, *17*(2), 109–135.

Bell, T., Logan, D., & Friedman, T. (2008). *Key issues for establishing information governance policies, processes and organization*.

Benbasat, I., David, & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly*, *11*(3), 369–386. https://doi.org/10.2307/248684

Benbasat, I., & Zmud, R. W. (1999). Empirical Research in Information Systems : The Practice of Relevance. *MIS Quarterly*, *23*(1), 3–16.

Bhaskar, R. (1978). *A Realist Theory of Science* (2nd ed.). Harvester.

Black, A., & van Nederpelt, P. (2020). *Dimensions of Data Quality (DDQ)*.

Black, S. A., & Porter, L. J. (1996). Identification of the Critical Factors of TQM. *Decision Sciences*, *27*(1), 1–21.

Brown, S. R. (1993). A Primer on Q Methodology. *Operant Subjectivity*, *16*(3/4), 91–138.

Caldeira, M. (2000). Critical Realism: A philosophical perspective for case study research in social sciences. *Episteme*, *2*(5–6), 73–88.

Côrte-Real, N., Ruivo, P., & Oliveira, T. (2020). Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value? *Information and Management*, *57*(1). https://doi.org/10.1016/j.im.2019.01.003

DAMA. (2008). *The DAMA Dictionary of Data Management* (M. Mosley, Ed.; 1st ed.). Technics Publications, LLC.

DAMA. (2017). *DAMA - DMBOK Data Management Body of Knowledge, 2nd Edition* (2nd ed.). Technics Publications.

DAMA International. (2009). *The DAMA Guide to the Data Management Body of Knowledge* (M. Mosley, M. Brackett, & S. Earley, Eds.). Technics Publications, LLC.

Dane, F. (1990). *Research Methods*. Brooks/Cole Publishing Company.

Daniel, D. (1961). Management information crisis. *Harvard Business Review*, *39*(5), 111–121. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:MANAGEMENT+INFORMATION+CRISIS#0

Davenport, T. H., & Patil, D. J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*, *October 2012*, 70–77.

Davenport, T. H., & Tiwari, P. (2024). Is Your Company's Data Ready for Generative AI? *Harvard Business Review*, 1–7.

Davenport, T., & Prusak, L. (1998). Working Knowledge: How Organizations Manage What They Know. *ACM Ubiquity*. http://www.acm.org/ubiquity/book/t_davenport_1.html

Drucker, P. F. (1989). *The New Realities*. Harper & Row.

Drucker, P. F. (2003). *A Functioning Society*. Transaction Publishers.

EFQM. (2013). *An Overview of the EFQM Excelence Model*.

English, L. P. (1999). *Improving Data warehouse and Business Information Quality*. John Wiley & Sons, Inc.

Fereday, J., & Muir-cochrane, E. (2006). Demonstrating Rigor Using Thematic Analysis : A Hybrid Approach of Inductive and Deductive Coding and Theme Development. *International Journal of Qualitative Methods*, *5*(1), 80–92.

Gabr, M., Helmy, Y. M., & Elzanfaly, D. S. (2021). Data Quality Dimensions, Metrics, and Improvement Techniques. *Future Computing and Informatics Journal*, *6*(1). https://digitalcommons.aaru.edu.jo/fcij/vol6/iss1/3

Gao, J., Xie, C., & Tao, C. (2016). Big data validation and quality assurance - Issuses, challenges, and needs. *Proceedings - 2016 IEEE Symposium on Service-Oriented System Engineering, SOSE 2016*, 433–441. https://doi.org/10.1109/SOSE.2016.63

Gartner. (n.d.). *Data Quality Tools*. Https://Www.Gartner.Com/En/Information-Technology/Glossary/Data-Quality-Tools.

Golafshani, N. (2003). Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report*, *8*(4), 597–606.

Gregor, S. (2006). The Nature of Theory in Information Systems. *MIS Quarterly*, *30*(3), 611–642.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337–355. https://doi.org/10.25300/MISQ/2013/37.2.01

Hazen, B. T., Boone, C. a., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, *154*, 72–80. https://doi.org/10.1016/j.ijpe.2014.04.018

Hevner, A., & Chatterjee, S. (2010). *Design Research in Information Systems* (Vol. 22). Springer US. https://doi.org/10.1007/978-1-4419-5653-8

Hevner, A. R. (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, *19*(2), 87–92. https://doi.org/http://aisel.aisnet.org/sjis/vol19/iss2/4

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.1007/BF01205282

Hietschold, N., Reinhardt, R., & Gurtner, S. (2014). Measuring critical success factors of TQM implementation successfully – a systematic literature review. *International Journal of Production Research*, *July*, 1–19. https://doi.org/10.1080/00207543.2014.918288

Ho, G. W. K. (2017). Examining Perceptions and Attitudes: A Review of Likert-Type Scales Versus Q-Methodology. *Western Journal of Nursing Research*, *39*(5), 674–689. https://doi.org/10.1177/0193945916661302

Hoffer, J. A., Ramesh, V., & Topi, H. (2016). *Modern Database Management* (12th ed.). Pearson Education Limited.

IBM. (n.d.). *What is data lineage?* Https://Www.Ibm.Com/Topics/Data-Lineage.

IBM Data Governance Council. (2007). *The IBM Data Governance Council Maturity Model : Building a roadmap for effective data governance* (Issue October).

Iivari, J. (2007). A Paradigmatic Analysis of Information Systems as a Design Science. *Scandinavian Journal of Information Systems*, *19*(2), 39–64.

ISO 8000-8:2015, Pub. L. No. ISO 8000-8 (2015). www.iso.org

ISO 8000-8:2015(E). www.iso.org

ISO 9001:2015 (2015).

ISO Standard 8000 Part (2) Version (1) - Data Quality : Vocabulary, 2017 (2017).

ISO/IEC 25012, 2008 (2008).

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, *2*, 59–64. https://doi.org/10.1016/j.bdr.2015.01.006

Kahn, B., Strong, D., & Wang, R. (2002). Information quality benchmarks: product and service performance. *Communications of the ACM*, *45*(4), 184–192. https://doi.org/10.1145/505999.506007

Karkošková, S. (2023). Data Governance Model To Enhance Data Quality In Financial Institutions. *Information Systems Management*, *40*(1), 90–110. https://doi.org/10.1080/10580530.2022.2042628

Kaur, N., & Sood, S. K. (2017). Dynamic resource allocation for big data streams based on data characteristics (5Vs). *International Journal of Network Management*, *27*(4). https://doi.org/10.1002/nem.1978

Khatri, V., & Brown, C. V. (2010). Designing Data Governance. *Communications of the ACM*, *53*(1), 148–152.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, *3*(3), 262–267. https://doi.org/10.1177/2043820613513388

Kivunja, C., & Kuyini, A. B. (2017). Understanding and Applying Research Paradigms in Educational Contexts. *International Journal of Higher Education*, *6*(5), 26. https://doi.org/10.5430/ijhe.v6n5p26

Knight, M. (2021). *What Is a Chief Data Officer (CDO)?* Dataversity. https://www.dataversity.net/chief-data-officer-cdo/#

Kobielus, J. (2022). *2022 State of Data Quality*.

Kobielus, J. (2024). *2024 State of Data Quality*.

Laney, D. (2001, February 6). *Meta Group*. 3D Data Management: Controlling Data Volume, Velocity, and Variety. http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-%0AData-Volume-Velocity-and-Variety.pdf%0A

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big Data, Analytics and the Path from Insights to Value. *MIT Sloan Management Review*, *52*(2), 21–32.

Linstone, H. A. (1981). The multiple perspective concept: With applications to technology assessment and other decision areas. *Technological Forecasting and Social Change*, *20*(4), 275–325.

Linstone, H. A., & Turoff, M. (2002). *The Delphi method - Techniques and Applications*. https://doi.org/10.1007/s00256-011-1145-z

Liu, C., Peng, G., Kong, Y., Li, S., & Chen, S. (2021). Data quality affecting big data analytics in smart factories: Research themes, issues and methods. *Symmetry*, *13*(8). https://doi.org/10.3390/sym13081440

Loshin, D. (2011). *Evaluating the Business Impacts of Poor Data Quality*. www.knowledge-integrity.com

Lucas, A. (2010). Towards Corporate Data Quality Management. *Portuguese Journal of Management Studies*, *XV*(2), 173–196.

Lucas, A., & Palma-dos-Reis, A. (2008). Reflexões sobre o Desafio da Relevância na Investigação em Sistemas de Informação Resumo. *Atas da 8ª Conferência Da Associação Portuguesa de Sistemas de Informação*, 1–13.

Madnick, S. E., Wang, R. Y., Lee, Y. W., & Zhu, H. (2009). *Overview and Framework for Data and Information Quality Research*. *1*(1), 2:1-2:22. https://doi.org/10.1145/1515693.1516680.http

March, S., & Smith, G. (1995). Design and Natural Science Research on Information Technology. *Decision Support Systems*, *15*, 251–266. https://doi.org/10.1016/0167-9236(94)00041-2

Mashey, J. R. (1999). *Big Data and the next wave of infraStress - Problems, Solutions, Opportunities*. Proceedings of the 1999 USENIX Annual Technical Conference. http://static.usenix.org/event/usenix99/invited_talks/mashey.pdf

Merton, R. K. (1947). Selected Problems of Field Work in the Planned Community. *American Sociological Review*, *12*(3), 304–312.

Mingers, J. (2001). Combining IS Research Methods: Towards a Pluralist Methodology. *Information Systems Research*, *12*(3), 240–259. https://doi.org/10.1007/s00415-007-0569-9

NIST. (2017). *Baldrige Performance Excellence Framework*.

OECD. (2015). *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation* (The Measurement of Scientific, Technological and Innovation Activities). OECD. https://doi.org/10.1787/9789264239012-en

Oliveira, P., Rodrigues, F., Henriques, P., & Galhardas, H. (2005). A Taxonomy of Data Quality Problems. *2nd Int. Workshop on Data and Information Quality*, 219–233.

Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., & Sinz, E. J. (2011). Memorandum on design-oriented information systems research. *European Journal of Information Systems*, *20*(1), 7–10. https://doi.org/10.1057/ejis.2010.55

Patton, M. Q. (2002). *Qualitative evaluation and research methods* (3rd ed.). Sage Publications, Inc.

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

Plotkin, D. (2021). *Data Stewardship An Actionable Guide to Effective Data Management and Data Governance* (2nd ed.). Elsevier Inc.

Porter, L. J., & Parker, A. J. (1993). Total quality management — the critical success factors. *Total Quality Management*, *4*(I), 13–23.

Principles for Effective Risk Data Aggregation and Risk Reporting, 28 (2013). https://doi.org/92-9197-138-3

Ramasamy, A., & Chowdhury, S. (2020). Big Data Quality Dimensions: A Systematic Literature Review. *Journal of Information Systems and Technology Management*. https://doi.org/10.4301/s1807-1775202017003

Redman, T. (2017). Seizing Opportunity in Data Quality. *MIT Sloan Management Review*.

Redman, T. C. (1996). *Data Quality for the Information Age*. Artech House, Inc.

Redman, T. C. (2016). Bad Data Costs the U.S. $3 Trillion Per Year. *Harvard Business Review*.

Rockart, J. F. (1979). Chief executives define their own data needs. *Harvard Business Review, 39*(5), 81–94.

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forescasting*, *15*, 353–375.

Royal Mail, & DataIQ. (2016). *How better customer data drives marketing performance and business growth*.

Sakpal, M. (2021, July 14). *D&A leaders must take pragmatic and targeted actions to improve their enterprise data quality if they want to accelerate their organizations' digital transformation*. Https://Www.Gartner.Com/Smarterwithgartner/How-to-Improve-Your-Data-Quality.

Santos, M. P. da C. dos. (2015). *Fatores Críticos de Sucesso na Gestão da Qualidade dos Dados*.

Santos, L., & Amaral, L. (2004). Estudos Delphi com Q-Sort sobre a web – A sua utilização em Sistemas de Informação. *Proceedings of the Conferência Da Associação Portuguesa Dos Sistemas de Informação*.

Santos, M. P., & Lucas, A. (2019). Identifying Critical Success Factors for Data Quality Management through a Delphi Study. *International Journal of Computer and Information Engineering*, *13*(8), 437–442.

Saraph, J. V., Benson, P. G., & Schroeder, R. G. (1989). An Instrument for Measuring the Critical Success Factors of Quality Management. *Decision Sciences*, *20*(4), 810–829.

Saunders, M. N. K., Lewis, P., & Thornhill, A. (2019). Understanding research philosophy and approaches to theory development. In *Research methods for business students* (8th ed.). Pearson Education Limited.

Schmidt, R. C. (1997). Managing Delphi surveys using nonparametric statistical techniques. *Decision Sciences*, *28*(3), 763–774. https://doi.org/10.1111/j.1540-5915.1997.tb01330.x

Simon, H. A. (1988). The Science of Design: Creating the Artificial. *Design Issues*, *4*(1/2), 67–82.

Skulmoski, G. J., & Hartman, F. T. (2007). The Delphi Method for Graduate Research. *Journal of Information Technology Education*, *6*, 1–21.

Sobreperez, P. (2008). Using Plenary Focus Groups in Information Systems Research : More than a Collection of Interviews. *The Electronic Journal of Business Research Methods*, *6*(2), 181–188.

Stephenson, W. (1953). *The Study of Behavior: Q-technique and its Methodology*. University of Chicago Press.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Sage Publications, Inc.

Sveiby, K. (1997). *The new organisational wealth - Managing and measuring Knowledge-Based Assets*. Berret – Koehler.

Tee, S. W., Bowen, P. L., Doyle, P., & Rohde, F. H. (2007). Factors influencing organizations to improve data quality in their information systems. *Accounting and Finance*, *47*(2007), 335–355.

Vaus, D. A. de. (2001). The Context of Design. In *Research Design in Social Research* (pp. 1–16). Sage Publications, Ltd.

Wahyuni, D. (2012). The Research Design Maze: Understanding Paradigms, Cases, Methods and Methodologies. *Journal of Applied Management Accounting Research*, *10*(1), 69–80.

Wand, Y., & Wang, R. Y. (1996). Anchoring Data Quality Dimensions in Ontological Foundations. *Communications of the ACM*, *39*(11), 86–95.

Wand, Y. ;, & Weber, R. (2004). Reflection: Ontology in Information Systems. *Journal of Database Management*, *15*(2), iii–vi.

Wang, R. Y. (1998). A Product Perspective onTotal Data Quality Management. *Communications of the Acm*, *41*(2).

Wang, R. Y., & Strong, D. M. (1993). Beyond Accuracy : What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*(4), 5–34.

Weber, K., Otto, B., & Osterle, H. (2009). One Size Does Not Fit All—A Contingency Approach to Data Governance. *ACM Journal OfData and Information Quality*, *1*(1). https://doi.org/10.1145/1515693.1515696.http

Wende, K. (2007). A model for data governance - Organising accountabilities for data quality management. *ACIS 2007 Proceedings - 18th Australasian Conference on Information Systems*, 416–425.

Williams, T. L., Becker, D. K., Robinson, C., Redman, T. C., & Talburt, J. R. (2015). Measuring Sociocultural Factors of Success in Information Quality Projects. *Proceedings of the 20th International Conference on Information Quality (ICIQ-2015)*, 48–69.

Xu, H. (2003). Would Organization Size Matter for Data Quality. *Proceedings of the Eighth International Conference on Information Quality (ICIQ-03)*.

Xu, H. (2015). What Are the Most Important Factors for Accounting Information Quality and Their Impact on AIS Data Quality Outcomes ? *ACM Journal of Data and Information Quality*, *5*(4), 14:1-14:22.

Xu, H., Koronios, A., & Brown, N. (2003). Managing Data Quality in Accounting Information Systems. In *IT-based management : challenges and solutions* (pp. 277–299). Idea Group Publishing.

Xu, H., & Lu, D. (2003). The Critical Success Factors for Data Qualty in Accounting Information System - Different Industries' Perspective. *Issues in Information Systems*, *4*, 762–768.

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

Yin, R. K. (2016). *Qualitative Research from Start to Finish*. The Guilford Press.

Zaletel, M., & Kralj, M. (2015). *Methodological guidelines and recommendations for efficient and rational governance of patient registries*.

Zmud, R. W. (1996). Editor' s Comments on Rigor and Relevancy. *MIS Quarterly*, *20*(3).

APPENDIX A: FOCUS GROUP REPORT

A Focus Group was held at a meeting room in ISEG, University of Lisbon. The duration of the meeting was 1 hour and 16 minutes and the 6 attendees were recognized academics and practitioners. (Belanger, 2012) prescribes that the number of participants in a focus group should be from 3 to 10, the most common number being 5 to 7, and the duration should be between 1 and 2 hours. I carried out the role of moderator.

I gratefully acknowledge the contributions of the participants who took part in this focus group meeting.

The focus group discussion was recorded, with the participants' permission.

Prior to the meeting, the participants were sent a list of the CSFs (Table XV) by email, and they were informed of the focus group's objective: to try to reduce the number of CSF. The meeting sought consensus in the answer to the following three questions:

- Is there any CSF that should not be considered as such?
- Are there some CSFs missing? Which ones?
- Is there any CSF that can be merged with another or others?

The moderator began the meeting defining the concept of CSFs. She then reminded the participants of the above questions to discuss.

Four participants agreed to add "data governance" to the list of CSFs, and one of them defined data governance as "set of key actions to ensure data compliance with organizational strategies".

Another participant considered that the definition of "organization for quality" corresponds to the concept of "data governance" and suggested that the CSF "teamwork" be included in "data governance". The same participant proposed that "physical environment", "storage management" and "nature of IS" should be considered as contingency factors, not CSFs for DQ. The remaining participants did not oppose those proposals.

One participant said there was a lot of intersection in the CSFs definitions and the whole group agreed with him. The moderator agreed with the group and asked the participants to help, as far as possible, to redefine the CSFs whose definitions intersect. This was partially achieved, as will be seen below.

One participant outlined the essential CSFs, from her point of view:

- Data quality should be in line with the strategic plan. This CSF was later included in the "Management Commitment and Leadership" CSF;
- DQ policies and standards;
- Monitoring and control of compliance with policies and standards;
- Input validations;
- Organizational culture with a focus on DQ;
- Top Management commitment to DQ;
- External data flows (included in CSF "Information Product Lifecycle Management"). She explained that this CSF gained great importance with the entry into force of the General Data Protection Regulation (GDPR), on May 28, 2018.This CSF has an enormous relevance to respond to external reporting.

Another participant pointed out that, from his point of view, the most relevant CSFs are:

- Data Quality Assessment/Monitoring;
- Stakeholder identification (included in the CSF "Data Quality Requirements Management");
- Management Commitment and Leadership;
- Continuous Data Quality Management Improvement;
- Focus on Data Customer Satisfaction;
- Information Architecture Management, considered as the architecture of the IS ecosystem, or its geography, is relevant to the type of DQ initiatives. The architecture of IS ecosystem is related to the flows of information, as well as the ownership of the data in each system.

The same participant suggested adding a CSF which he termed "company maturity with respect to data processing", which was seconded by another participant. The moderator argued that the "maturity" should not be considered a CSF, because it corresponds to the state of the company, prior to any DQ initiative, and therefore cannot be considered as "an area that must be given special and continual attention to succeed in DQ initiatives". The participant presented an example of the applicability of the CSF to a company that has just started their activity. The moderator argued that it is a very special case, and the proposed CSF should not be considered.

Two participants suggested the CSF “strategic data quality planning” could be considered as “Data quality should be in line with the strategic plan” and one participant suggested that it could be included in the CSF “management commitment and leadership”.

It was proposed that the CSF “understanding of the information systems and DQ” be renamed as “understanding of the information systems and the relevance of DQ”.

After some discussion, the participants reached a consensus to include the CSFs "continuous improvement" and “DQ controls/input controls” in the CSF "continuous information quality management improvement". One participant suggested including in the description of "continuous information quality management improvement " the following: “set of actions we must take to improve data quality” and gave an example of a change in a data collection screen to facilitate the insertion of certain data.

The group agreed to change the name of the CSF “customer focus and satisfaction” to “focus on data customer satisfaction”, and some discussion about the concept of data customer was generated. The group came to the consensus that there are two types of data customer: the clients (external customers) and the so-called users, who are the internal customers.

After some discussion the group agreed to include micro changes, such as the change of an attribute domain, in the CSF "management of changes”. One participant proposed, and it was accepted, to change the definition of “management of changes” to include the following “active inclusion of the DQ requirements updating in the context of management of changes”.

One participant proposed to change the name of the CSF “risk management” to “DQ risk management”, which was accepted.

After the focus group meeting finished, I considered that like other CSFs that were considered contingency factors, the CSF "employee relations” should also be considered a contingency one.

As a result of the meeting, I reorganized the list of CSFs, changed some of their names and modified some of the definitions as agreed at the focus group.

In the final review of the CSFs list, I decided that the definition of IQ-KPIs is better included in CSF “Information Quality Assessment/Monitoring” and therefore removed it from CSF “Continuous Information Quality Management Improvement” and included it in “Information Quality Assessment/Monitoring”.

The report of the focus group meeting, as well as the amended list of CSFs (Table XXI), was sent to the participants, requesting their feedback.

APPENDIX B: CODE RELATIONS

Code System	Data Governance	Management Commitment and Leadership	Training	Data Quality Requirements Management	Supplier Partnership	Data Product Lifecycle Management	Data Quality Assessment/ Monitoring
Data Governance	0	2	3	2	2	5	5
Management Commitment and Leadership	2	0	1	0	0	2	1
Training	3	1	0	1	2	2	3
Data Quality Requirements Management	2	0	1	0	1	1	1
Supplier Partnership	2	0	2	1	0	1	2
Data Product Lifecycle Management	5	2	2	1	1	0	2
Data Quality Assessment/Monitoring	5	1	3	1	2	2	0
Focus on Data Customer Satisfaction	6	0	2	2	2	3	4
Continuous Data Quality Management Improvement	7	1	3	2	2	4	5
Culture and communication	6	1	2	2	1	3	4
Data Architecture Management	9	2	3	2	2	5	5
Data Security Management	1	0	1	1	1	0	1
DQ Risk Management	6	2	3	1	2	4	3
Understanding of the IS and the relevance of DQ	0	0	0	0	0	0	0
Personnel Competency	4	1	2	0	1	3	2
DQ Policies and Standards	5	1	0	1	0	3	2
Management of Changes	2	0	0	0	0	1	1
Audits and Reviews	1	0	1	1	1	0	1
Evaluate cost/benefit trade-offs	2	0	1	0	1	2	1
Business Processes Change	1	0	0	1	0	1	0
Information Catalog	8	2	2	2	1	4	4
DQ Tools	7	1	2	2	1	3	4
Sponsorship	5	1	2	1	1	3	3
Management of User Expectations	1	0	0	0	0	0	1
Business Model Definition	1	0	0	0	0	0	1

Code System	Focus on Data Customer Satisfaction	Continuous Data Quality Management Improvement	Culture and communication	Data Architecture Management	Data Security Management	DQ Risk Management	Understanding of the IS and the relevance of DQ
Data Governance	6	7	6	9	1	6	0
Management Commitment and Leadership	0	1	1	2	0	2	0
Training	2	3	2	3	1	3	0
Data Quality Requirements Management	2	2	2	2	1	1	0
Supplier Partnership	2	2	1	2	1	2	0
Data Product Lifecycle Management	3	4	3	5	0	4	0
Data Quality Assessment/Monitoring	4	5	4	5	1	3	0
Focus on Data Customer Satisfaction	0	6	5	6	1	3	0
Continuous Data Quality Management Improvement	6	0	6	7	1	4	1
Culture and communication	5	6	0	6	1	3	0
Data Architecture Management	6	7	6	0	1	6	0
Data Security Management	1	1	1	1	0	1	0
DQ Risk Management	3	4	3	6	1	0	0
Understanding of the IS and the relevance of DQ	0	1	0	0	0	0	0
Personnel Competency	2	3	2	4	0	4	0
DQ Policies and Standards	4	4	4	5	0	2	0
Management of Changes	2	2	2	2	0	1	0
Audits and Reviews	1	1	1	1	1	1	0
Evaluate cost/benefit trade-offs	2	2	1	2	0	2	0
Business Processes Change	1	1	1	1	0	0	0
Information Catalog	5	7	6	8	1	5	1
DQ Tools	5	6	6	7	1	4	0
Sponsorship	3	5	3	5	0	3	1
Management of User Expectations	1	1	1	1	0	0	0
Business Model Definition	1	1	1	1	0	0	0

Code System	Personnel Competency	DQ Policies and Standards	Management of Changes	Audits and Reviews	Evaluate cost/benefit trade-offs	Business Processes Change	Information Catalog
Data Governance	4	5	2	1	2	1	8
Management Commitment and Leadership	1	1	0	0	0	0	2
Training	2	0	0	1	1	0	2
Data Quality Requirements Management	0	1	0	1	0	1	2
Supplier Partnership	1	0	0	1	1	0	1
Data Product Lifecycle Management	3	3	1	0	2	1	4
Data Quality Assessment/Monitoring	2	2	1	1	1	0	4
Focus on Data Customer Satisfaction	2	4	2	1	2	1	5
Continuous Data Quality Management Improvement	3	4	2	1	2	1	7
Culture and communication	2	4	2	1	1	1	6
Data Architecture Management	4	5	2	1	2	1	8
Data Security Management	0	0	0	1	0	0	1
DQ Risk Management	4	2	1	1	2	0	5
Understanding of the IS and the relevance of DQ	0	0	0	0	0	0	1
Personnel Competency	0	1	1	0	2	0	3
DQ Policies and Standards	1	0	2	0	1	1	5
Management of Changes	1	2	0	0	1	0	2
Audits and Reviews	0	0	0	0	0	0	1
Evaluate cost/benefit trade-offs	2	1	1	0	0	0	1
Business Processes Change	0	1	0	0	0	0	1
Information Catalog	3	5	2	1	1	1	0
DQ Tools	3	4	2	1	1	1	7
Sponsorship	3	2	0	0	1	1	5
Management of User Expectations	0	1	1	0	0	0	1
Business Model Definition	0	1	1	0	0	0	1

Code System	DQ Tools	Sponsorship	Management of User Expectations	Business Model Definition
Data Governance	7	5	1	1
Management Commitment and Leadership	1	1	0	0
Training	2	2	0	0
Data Quality Requirements Management	2	1	0	0
Supplier Partnership	1	1	0	0
Data Product Lifecycle Management	3	3	0	0
Data Quality Assessment/Monitoring	4	3	1	1
Focus on Data Customer Satisfaction	5	3	1	1
Continuous Data Quality Management Improvement	6	5	1	1
Culture and communication	6	3	1	1
Data Architecture Management	7	5	1	1
Data Security Management	1	0	0	0
DQ Risk Management	4	3	0	0
Understanding of the IS and the relevance of DQ	0	1	0	0
Personnel Competency	3	3	0	0
DQ Policies and Standards	4	2	1	1
Management of Changes	2	0	1	1
Audits and Reviews	1	0	0	0
Evaluate cost/benefit trade-offs	1	1	0	0
Business Processes Change	1	1	0	0
Information Catalog	7	5	1	1
DQ Tools	0	4	1	1
Sponsorship	4	0	0	0
Management of User Expectations	1	0	0	1
Business Model Definition	1	0	1	0

APPENDIX C: CODEBOOK

05/05/2024

Code System

1 Data Governance	7
1.1 Data Stewardship	10
2 Management Commitment and Leadership	4
3 Training	3
4 Data Quality Requirements Management	3
5 Supplier Partnership	3
6 Data Product Lifecycle Management	8
7 Data Quality Assessment/Monitoring	8
8 Focus on Data Customer Satisfaction	9
9 Continuous Data Quality Management Improvement	18
9.1 Input Data Validations	11
10 Culture and communication	11
11 Data Architecture Management	10
11.1 Data Ownership	10
12 Data Security Management	1
13 DQ Risk Management	11
14 Understanding of the IS and the relevance of DQ	1
15 Personnel Competency	8
16 DQ Policies and Standards	6
17 Management of Changes	2
18 Audits and Reviews	1
19 Evaluate cost/benefit trade-offs	7
20 Business Processes Change	2
21 Information Catalog	18
22 DQ Tools	11
23 Sponsorship	9
24 Management of User Expectations	1
25 Business Model Definition	2

1 Data Governance

This entails a series of important steps to guarantee that data are compliant with organizational strategies (Focus Group). It establishes a proper organizational structure for the generation of high-quality data. It should specify who is responsible for the DQ: appoint data stewards and a data champion (Santos, 2015); appoint an expert or a group of experts as DQ managers. Promote teamwork between business and IT people, as a key to improve data quality (Xu et al., 2003) (Focus Group).

1.1 Data Stewardship

Communities of Interest focused on one or more specific subject-areas or projects, collaborating or consulting with project teams on data definitions and data management standards related to the focus. Consists of business and technical data stewards and data analysts DAMA (2017)

2 Management Commitment and Leadership

Top management must establish a solid foundation of clear data quality values and policies, as well as supply the necessary resources (Hietschold et al., 2014).

"Data quality should be in line with the strategic plan" (Focus Group). To achieve consistent and long-term excellence, companies must incorporate data quality into their organizational strategy (Hietschold et al., 2014).

3 Training

Employee engagement and empowerment require knowledge of data quality concepts, methods, and tools (Hietschold et al., 2014). Training needs should be recognized and documented, and training workshops should be held on a regular basis. Mentoring programs should ensure on-the-job professional development in addition to formal training (Baskarada & Koronios, 2014).

4 Data Quality Requirements Management

All relevant stakeholders should be identified, and their requirements collected and modelled. (Baskarada & Koronios, 2014).

5 Supplier Partnership

Data supplier quality management entails establishing a successful data quality management relationship with raw data suppliers. There are two key components:

1. To come to an agreement on the acceptable level of raw data quality, including requirements for availability, currency, accuracy, and completeness; 2. To provide data suppliers with regular data quality reports and technical assistance (Xu et al., 2003).

6 Data Product Lifecycle Management

Managing information as a product as well as effectively managing the information processes (life cycles of critical information products) is important for effective data quality management. One of the aspects of this CSF includes identifying and documenting the data flow within the organization as well as between the organization and any external parties (i.e., information product supply chain management) (Baskarada & Koronios, 2014). Clarity of process ownership (process owners), boundaries, and steps are very important (Saraph et al., 1989).

7 Data Quality Assessment/Monitoring

Prior to attempting any DQ improvements, the existing status of DQ must be assessed, and qualitative and quantitative metrics must be developed and used (DQ-KPI) (Baskarada & Koronios, 2014).

It should be possible to access most of the data quality dimensions using profiling tools. DQM metrics or Key Performance Indicators (DQ-KPIs) should be defined qualitatively and quantitatively, and then utilized to continually analyze the efficacy of corporate DQM activities (Baskarada & Koronios, 2014) (Focus Group). DQ should be evaluated at regular intervals using the same data profiling tools. In addition, policy and standard compliance should be monitored.

8 Focus on Data Customer Satisfaction

This means concentrating on the needs and quality requirements of data clients. It should allow data clients to actively participate in ensuring and improving data quality (Xu et al., 2003). Clients (external customers) and internal customers are both referred to as "data customers."

9 Continuous Data Quality Management Improvement

“Set of actions we must take to improve data quality” (Focus Group). There is a need for continuous and consistent data quality improvement, materialized as a set of actions that must be taken to improve data quality, such as input validations and human data quality controls (Xu et al., 2003).

9.1 Input Data Validations

The practice of comparing input that the program receives to a standard that is specified within the application is known as input validation.

10 Culture and communication

Encouragement of a data quality improvement culture within the organization (S. A. Black & Porter, 1996). Communication is considered a two-way process, with channels for feedback available. Communication is viewed as an ongoing process, with attention given to approaches of reinforcing the concepts in the future (Porter & Parker, 1993).

11 Data Architecture Management

Architecture of the IS ecosystem, or its geography, is relevant to the type of DQ initiatives. The architecture of the IS ecosystem should be described, namely the flows of information should be depicted, and the ownership of the data in each system should be identified (Focus Group).

11.1 Data Ownership

The term "data ownership" describes the responsibility for information as well as its possession. Both control and power are implied by ownership. Information control encompasses not only the capacity to access, create, modify, package, profit from, transfer, or delete data, but also the authority to provide other people with access rights to that data.

12 Data Security Management

Access security is a critical DQ dimension, and data security management necessitates the implementation of effective access controls that ensure that all users are properly authenticated and authorized with the bare minimum of privileges. IS developers, for example, should not be given access to the production environment. Audit trails (logs of users' activities on the IS) must also be analyzed (for example, for exceptions) and reviewed on a regular basis (Baskarada & Koronios, 2014).

13 DQ Risk Management

Risk management can be defined as the awareness of the implications of poor DQ and the level of commitment to reducing them (Xu et al., 2003). DQ risks to company objectives (such as financial, reputational, and regulatory risks) must be diagnosed, documented, assessed, categorized, prioritized, and mitigated/controlled. Effective DQ Risk Management should allow organizations to concentrate their DQM efforts on the most critical information products, resulting in increased DQM efficiency and effectiveness (Baskarada & Koronios, 2014).

14 Understanding the IS and the relevance of DQ

It is important to understand how the information systems work (technical competence) and IT personnel and data customers need to understand the importance of data quality (Xu et al., 2003).

15 Personnel Competency

The ability of employees in charge of IS is especially important; for example, exceptionally talented and knowledgeable personnel in both technical and business areas are required (Xu et al., 2003).

16 DQ Policies and Standards

Simple, relevant, and consistent data quality policies and standards must be in place inside the organization. There are two primary parts:

1. Establishing specified and relevant data quality policies and standards;

2. Implementing and enforcing policies and standards (Xu et al., 2003).

17 Management of Changes

"Active inclusion of the DQ requirements updating in the context of management of changes" (Focus Group). DQ requirements, which can be internal or external, should be included and consistently updated in the process of management of changes. Internal changes include structural changes, such as organizational restructuring, as well as micro changes, such as the change of an attribute domain. External changes include things such as government regulations, technology, economy, and market changes (Xu et al., 2003).

18 Audits and Reviews

Independent internal and external audits of the systems and data quality should be conducted to ensure that appropriate controls are in place, and data quality reviews should be conducted regularly (Xu et al., 2003).

19 Evaluate cost/benefit trade-offs

It's critical to evaluate the costs of poor DQ and corresponding improvement activities, as well as any potential benefits or cost savings that may result from any process enhancements, before moving forward with any process improvements (Baskarada & Koronios, 2014).

20 Business Processes Change

There should be flexibility to change business processes. DQ processes close to the customer should be used.

21 Information Catalog

An information catalog that supports efficient metadata management should be created. It must contain Technical definitions and Business Concepts to be queried by some users. For Business Concepts it can be a dictionary of terms (organizational concepts and their synonyms).

22 DQ Tools

Data quality tools are the processes and technologies for identifying, understanding and correcting flaws in data that support effective information governance across operational business processes and decision making. The packaged tools available include a range of critical functions, such as profiling, parsing, standardization, cleansing, matching, enrichment and monitoring (Gartner Glossary https://www.gartner.com/en/information-technology/glossary/data-quality-tools).

More recently, corporations are using AI-driven DQM automation, which according to TDWI (Kobielus, 2022), refers to the use of embedded machine learning models to handle one or more DQM functions reliably, repeatedly, and accurately without needing direct human oversight and assistance.

23 Sponsorship

Someone at C-Level that has a vision of the importance of information assets quality.

24 Management of User Expectations

Avoid False expectations.

25 Business Model Definition

A well-defined business model allows us to identify what needs to be improved in the DQ.