Universiteit Utrecht

Become mature before it's too late

Developing a 'Data Quality Management Maturity Model' (DQM3) based on Critical Success Factors

Thesis report

Master of Science - Business Informatics
Department of Information and Computing Sciences
Faculty of Science - Utrecht University

Author

T. (Tim) Mathijsen, BSc.

5663458

t.mathijsen@students.uu.nl

Utrecht University

Project Supervisor

Drs. N.A. (Nico) Brand

n.a.brand.uu.nl

Utrecht University

Second Examiner

Dr. M.R. (Marco) Spruit

m.r.spruit@uu.nl

Utrecht University

Daily Supervisor

A. (Ari) Sadik, MSc.

ari.sadik@achmea.nl

Achmea IT

Secondary Daily Supervisor

S. (Sven) Huijsmans

sven.huijsmans@achmea.nl

Achmea IT

Abstract

Aims. In existing scientific literature, there is no maturity model for the data quality management (DQM) domain that provides sufficient supporting materials for an organization to evaluate their current status of DQM capability without being assisted by third-parties or certified professional (i.e., performing a self-assessment) and includes all Critical Success Factors (CSFs) for DQM. The CSFs for DQM are recently identified in two publications, while the current models were developed three to fifteen years ago. The existing DQM maturity models are also not applicable to measure the maturity level of a business chain. Therefore, the purpose of this study is twofold: (1) to develop a maturity model that allows an organization to perform a self-assessment in which the DQM maturity level of a business chain is measured, and (2) to apply and validate this maturity model by performing a pretest and a case study at Achmea.

Methodology. This study was structured according to the design cycle of Wieringa, supported by the development cycle of Mettler. Various research methods were used throughout this study. In the first phase of the design cycle, problem investigation, a literature study was conducted to confirm the gaps in the current scientific literature. In addition, an unstructured interview was performed and company reports of Achmea were studied to define the problem statement of this study. In the second phase, treatment design (supported by the 'define scope' and 'design model' tasks of the development cycle), the maturity model was developed based on an extensive literature study. In the third and final phase of the design cycle, treatment validation (supported by the 'evaluate design' task of the development cycle), a pretest and a case study was performed at Achmea to apply and validate the maturity model. Finally, the design mutability of the maturity model was contemplated in the final phase of the development cycle: reflect evolution.

Results. The treatment design phase led to the creation of the Data Quality Management Maturity Model (DQM3), consisting of two models: a three-layered domain reference model and an assessment model. The domain reference model is populated with identified and merged CSFs for (corporate) DQM and represents this domain in thirteen domain components and thirty-one domain sub-components. The assessment model consists of an assessment instrument (questionnaire implemented in Qualtrics Survey Software), five maturity levels, and defines how these maturity levels are assigned to the components of the domain reference model.

Conclusion. This study shows that the designed maturity model performs well in practice: both in the pretest and the case study. Only some minor improvements were made to the formulation of some questions. No additions have been made to the domain reference model. Future research is needed to refine and validate the domain reference model and assessment model of the DQM3 to demonstrate generalizability in other industries and organizations of various sizes.

Acknowledgements

The past eight months have been devoted to applying the knowledge and skills I gained during my bachelor and master study in this thesis project. It was an exciting and challenging period, mainly because of the corona crisis that started in the middle of the second phase. In the end, I managed to perform my research as I defined it in my research proposal. I could not accomplish this on my own, so I would like to take this opportunity to thank some of the people who helped and supported me during this process.

First of all, I would like to thank my project supervisor Nico Brand. He was also my first supervisor when finalizing my Information Science bachelor two years ago. Nico was always ready to provide useful feedback so that I could take my thesis report to a higher level. Although we hardly saw each other in personal meetings due to the corona crisis, I experienced our collaboration as very pleasant. In addition, I would like to thank Marco Spruit, the second examiner, for another pair of eyes on this thesis project.

I also want to thank my two daily supervisors at Achmea (Ari Sadik and Sven Huijsmans) for supporting me with useful feedback during my internship. If I had questions, they were always ready to answer them. Finally, I would like to thank my colleagues from the 'Datalogistiek' team for all the fun (also when working from home) and other colleagues at Achmea plus external participants from the Rabobank for participating in the pretest and case study. It would not have been possible without you!

I hope you enjoy reading this thesis!

Tim Mathijsen

Tilburg, July 13, 2020.

Table of Contents

Abstract 2

Acknowledgements 3

List of Figures 6

List of Tables 6

1. Introduction 7

1.1 Problem statement 7

1.2 Research goals and contributions 8

1.3 Thesis outline 8

2. Research Questions & Research Design 10

2.1 Research questions 10

2.2 Research design and research methods 10

2.2.1 Problem investigation 11

2.2.2 Treatment design and treatment validation 12

2.2.3 Conceptual framework 14

2.3 Literature research protocol 15

2.3.1 Search strategy 15

2.3.2 Selection criteria 15

2.3.3 Keywords 16

3. Literature Review: Background & Related Work 17

3.1 Data quality management & critical success factors 17

3.1.1 Data quality & data quality dimensions 17

3.1.2 Data quality management 19

3.1.3 Data quality management critical success factors: identification, ordering, and merging 20

3.1.4 Data quality management critical success factors: other studies 24

3.2 Maturity models & data quality management 25

3.2.1 Maturity model 25

3.2.2 Maturity model comparison framework 26

3.2.3 Data quality management maturity models: identification 29

3.2.4 Data quality management maturity models: comparison 30

4. Define Scope & Design Model (Treatment Design) 32

4.1 Define scope 32

4.2 Design model 33

4.2.1 Structure of the DQM3 34

4.2.2 Domain reference model of the DQM3 35

4.2.3 Assessment model of the DQM3 42

5. Evaluate Design (Treatment Validation) & Reflect Evolution 46

5.1 Evaluate design 46

5.1.1 Pretest of the assessment instrument 46

5.1.2 Case study at Achmea 48

5.2 Reflect evolution 55

6. Conclusion, Limitations & Future Research, & Reflection 56

6.1 Conclusion 56

6.2 Limitations and future research 58

6.3 Reflection on research goals and contributions 59

6.4 Personal reflection 59

References 60

Appendix 65

Appendix I – Interview about the DMM model of the CMMI Institute within Achmea (Dutch) 65

Appendix II – Critical Success Factors for (Corporate) Data Quality Management 67

A: CSFs for DQM identified by Santos & Lucas (2019) 67

B: CSFs for DQM identified by Lucas (2019) 69

Appendix III – Blueprint of the assessment instrument (version I) 71

Appendix IV –Blueprint of the assessment instrument (version II) 77

Appendix V – Blueprint of the assessment instrument (version III) 83

List of Figures

Figure 1 - Design Cycle (Adapted from Wieringa, 2014) 11
Figure 2 - Development cycle (Adapted from Mettler, 2011) 12
Figure 3 - Conceptual framework 14
Figure 4 - Types of maturity models (Adapted from Van Steenbergen et al., 2007) 27
Figure 5 - Components of the DQM3 (Adapted from Hüner, Ofner, & Otto, 2009) 34
Figure 6 - Components of the domain reference model 34
Figure 7 - Relation between the domain reference model and the final assessment model 45
Figure 8 - Chain of evidence (Adapted from Yin, 2003) 54

List of Tables

Table 1 - Guidelines for design science in information systems research (Adapted from Hevner et al., 2004) 11
Table 2 - Decision parameters of the development cycle (Adapted from Mettler, 2011) 12
Table 3 - Inclusion and exclusion criteria 15
Table 4 - Data quality dimensions and quality characteristics (Adapted from Jayawardene, Sadiq, & Indulska (2015)) 19
Table 5 - CSFs for DQM identified by Santos & Lucas (2019) 20
Table 6 - CSFs for corporate DQM identified by Lucas (2019) 21
Table 7 - Merged CSFs for DQM 23
Table 8 - Critical success factors for data quality management 23
Table 9 - Template to classify a maturity model (Adapted from Mettler et al. (2010) and Van Steenbergen et al. (2007)) 29
Table 10 - Classification of data quality management maturity models 31
Table 11 - Define scope decision parameters and characteristics 32
Table 12 - Design model decision parameters and characteristics 34
Table 13 - Domain reference model of the DQM3 (part I) 37
Table 14 - Domain reference model of the DQM3 (part II) 39
Table 15 - Domain reference model of the DQM3 (part III) 40
Table 16 - Domain reference model of the DQM3 42
Table 17 - Maturity levels, descriptors, and criteria 44
Table 18 - Example calculation of maturity scores and levels 45
Table 19 - Evaluate design parameters 46
Table 20 - Content difficulties and improvements made to the assessment instrument 47
Table 21 - Teams/departments identified for the case study 48
Table 22 - Missing components in the domain reference model 51
Table 23 - Revisions made to the DQM3 53
Table 24 - Case study validity and reliability concerns 53
Table 25 - Reflect evolution parameters 55
Table 26 - Critical success factors for data quality management (Adapted from Santos & Lucas, 2019) 68
Table 27 - Critical success factors for corporate data quality management (Adapted from Lucas, 2019) 70

1. Introduction

In today's world, the volume of data is increasing exponentially. Data is used in almost all the activities that companies perform and forms the basis for decisions on a strategic and operational level (Haug, Zachariassen, & Van Liemph, 2011). According to Davenport (2006), organizations that collect and use data strategically can gain a competitive advantage in their industry and sometimes even dominate it. However, poor data quality (DQ) has a negative influence on the efficiency of an organization (Haug, Zachariassen, & Van Liemph, 2011; Harexian, 2019). The typical impact of poor DQ is more inadequate decision making, increased organizational mistrust, and more challenging to set and execute strategy (Redman, 1998).

Achmea, an insurance company since 1811, is such an organization that tries to achieve this competitive advantage in the insurance industry. Achmea is a financial service provider that insures approximately ten million people (for healthcare, damage, and income[1]), is active in six different countries (including the Netherlands), has an annual gross premium revenue of almost €20 billion a year, and employs about 16,000 people[2].

Proper business operations within Achmea, as well as the business operations of their customers, strongly depend on accurate data from these ten million customers. However, in a complex organization like Achmea, data is subject to constant change (adding new fields or sources, deleting or changing existing data), which could negatively affect DQ. A recent case study conducted by Harexian (2019) showed that poor DQ causes a severe financial impact and still affects Interpolis (a subsidiary of Achmea) in several ways: rework by employees; unable to perform analysis; and email, payments, or debt collection sent to the wrong person or organization. Achmea's 'Datalogistiek' team, which is responsible for the data provision between source systems and internal/external customers, is also experiencing problems with poor DQ. This team is often forced to execute ad-hoc repair activities to improve DQ within the systems they manage.

To achieve this competitive advantage within the insurance industry, Achmea tries to tackle these issues with poor DQ by becoming a digital insurer that continuously and proactively manages the quality of its data. One of the initiatives within Achmea to achieve this mission is to evaluate and improve the current status of data management (DM) capability by using the Data Management Maturity (DMM) model of the Capability Maturity Model Integration (CMMI) Institute. A so-called 'quick scan' developed within Achmea helps senior managers with assessing their DM program based on the 25 processes areas of the DMM model. With this quick scan, Achmea can evaluate the DM capability of a business chain without being assisted by third-parties/certified professionals, i.e., performing a self-assessment.

1.1 Problem statement

The scientific literature on data quality management (DQM) emphasizes the difference between DM and DQM. According to Otto et al. (2007), DQM is a "quality-oriented DM, i.e., DM focusing on the collection, organization, storage, processing, and presentation of high-quality data." Since the mission of Achmea is to proactively and continuously manage the quality of its data, this suggests the need to evaluate and improve the current status of DQM capability in addition to or instead of the DMM model.

[1] https://www.achmea.nl/over-ons/organisatie

[2] https://www.achmea.nl/over-ons

However, as substantiated by the literature review in Chapter 3, there is no maturity model for the DQM domain that provides sufficient supporting materials to structure a self-assessment and includes all essential elements (also known as Critical Success Factors (CSFs)) for DQM. These CSFs should be applied within an organization to ensure competitive and successful performance (Huang, Wu, & Chen, 2012; Rockart, 1979), and are recently identified in two publications while the current maturity models were developed three to fifteen years ago. The existing DQM maturity models are also not applicable to measure the maturity level of a business chain. These limitations with the current maturity models and the business need of Achmea pose the following problem statement:

Problem statement

There exist no maturity model that allows an organization to perform a self-assessment in which the data quality management maturity level of a business chain is measured.

1.2 Research goals and contributions

The goal of this research is twofold. The first goal is to tackle the problem statement by developing a DQM maturity model that allows an organization to perform a self-assessment. This maturity model should provide organizations with the ability to measure the DQM maturity level of a business chain. If this research goal is achieved, the designed maturity model will contribute to the scientific body of knowledge and serve as a foundation for future research.

The second goal of this research is to apply and validate the designed DQM maturity model by performing a pretest and a case study at Achmea. This case study helps Achmea to evaluate the current status of DQM maturity in a given business chain. The results of the maturity assessment can then be used to serve as input for improving their maturity. This goal will provide the first practical contribution to this study. On the other hand, the new maturity model could potentially be used by other organizations to improve their DQM. Thus, the focus lies on providing supporting materials for performing a self-assessment, which is lacking in the existing DQM maturity models.

1.3 Thesis outline

The remaining of this thesis report is structured as follows:

- **Chapter 2 – Research Questions & Research Design**
  This chapter presents the research questions defined to solve the problem statement and to achieve the research goals, supported by a research design with corresponding research methods. The research design is supported by a conceptual framework that visualizes the coherence between the research questions, research methods, and deliverables of this study. Finally, the literature research protocol defines how relevant literature is identified and selected for this study.

- **Chapter 3 – Literature Review: Background & Related Work**
  In this chapter, a review of the existing scientific literature on DQM, CSFs for DQM, maturity models, and DQM maturity models is presented. This literature review confirms that the gap described in the problem statement exists and that the research approach defined in Chapter 2 is suitable and can lead to the scientific and practical contributions described above.

- **Chapter 4 – Define Scope & Design Model (Treatment Design)**

  This chapter presents the proposed solution to the problem statement: the 'Data Quality Management Maturity Model' (DQM3). The first section of this chapter is concerned with defining the scope of the model. The second and final section presents the three-layered domain reference model and the assessment model, which together form the DQM3.

- **Chapter 5 – Evaluate Model (Treatment Validation) & Reflect Evolution**

  This chapter presents the results of applying and validating the DQM3: the first section of this chapter discusses the results of a pretest and a case study conducted at Achmea. In the final section, the design mutability of the DQM3 is contemplated.

- **Chapter 6 – Conclusion, Limitations & Future Research, & Reflection**

  The final chapter of this thesis report consists of four sections: conclusion, limitations and future research, reflection on research goals and contributions, and personal reflection. The first part, conclusion, summarizes the answers to the sub-research questions and the main research question. Then, in the section 'limitations and future research', the research limitations and the possibilities for future research are discussed. In the last two parts of this final chapter, a reflection is given on the research goals and contributions, and a personal reflection on the intended learning objectives of this thesis project.

2. Research Questions & Research Design

This chapter takes a closer look at the research approach of this study. Section 2.1 presents the research questions defined to tackle and solve the problem statement. To answer these questions, section 2.2 defines a research design with corresponding research methods. This research design is supported by a conceptual framework in which the coherence between the research questions, research methods, and the deliverables are visualized. To complete this chapter, section 2.3 discusses the literature research protocol used for the literature review.

2.1 Research questions

The main research question and sub-questions to tackle the problem statement and achieve the research goals are:

How can an organization measure the data quality management maturity level of a business chain?

SQ1: *What are the critical success factors for data quality management?*

SQ2: *Which of the existing maturity models allows an organization to perform a self-assessment in which the data quality management maturity level of a business chain is measured?*

SQ3: *How could the identified critical success factors for data quality management be combined in a maturity model to measure the data quality management maturity level of a business chain?*

SQ4: *How is the designed maturity model used in practice?*

A detailed research design with corresponding research methods is needed to answer the research questions defined above. This research design will be discussed in the next section.

2.2 Research design and research methods

According to Wieringa (2014), 'design science' is the design and investigation of artifacts in context. These artifacts should be designed to interact with an identified problem context by improving something in that context. Because this study focuses on designing such an artifact (maturity model) that tries to improve a problem context (insight into the DQM maturity of a business chain), the problem statement defines a 'design problem'.

Since design problems are treated by executing the 'design cycle' of Wieringa (2014), this cycle is used as the primary research design of this study. The design cycle of Wieringa (2014), visualized in Figure 1, consists of three tasks: problem investigation, treatment design, and treatment validation. The first task, problem investigation, focuses on investigating what phenomena must be improved and why. Second, in the treatment design task, one or more artifacts are designed that could treat the problem identified in the previous task. Finally, the treatment validation task validates whether the proposed treatment treats the problem.

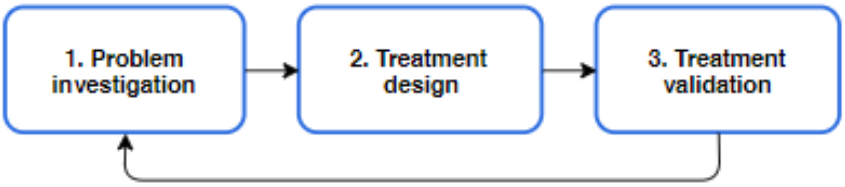


Figure 1 - Design Cycle (Adapted from Wieringa, 2014)

Hevner et al. (2004) define seven guidelines for design science in information systems research (Table 1). The research methods chosen to support the design cycle tasks are based on these seven guidelines to ensure that the designed maturity model will be scientifically valid.

Guidelines	Description
1. Design as an artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
2. Problem relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
3. Design evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
4. Research contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
5. Research rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
6. Design as a search process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
7. Communication of research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audience.

Table 1 - Guidelines for design science in information systems research (Adapted from Hevner et al., 2004)

2.2.1 Problem investigation

Appropriate research methods are selected to investigate the problem context thoroughly. According to Wieringa (2014), the most commonly used research methods for this problem investigation are surveys, observational case studies, single-case mechanism experiments, and statistical difference-making experiments. Because surveys can provide information about real-world phenomena, this research method is selected to obtain relevant information that contributes to identifying the problem and designing a suitable artifact.

A survey is conducted in the form of an unstructured interview with a Data Architect of Achmea (Appendix I). Also, relevant company reports of Achmea are studied to get a complete overview of the problem context. These surveys take the 'problem relevance' guideline of Hevner et al. (2004) into account. In addition to the surveys, a literature study is performed to investigate the concepts of DQM, CSFs for DQM, maturity models, and existing DQM maturity models, and to confirm the gap in the scientific literature. Section 2.3 describes the literature research protocol used for this literature study. The literature study takes the 'design as a search process' and 'research contribution' guidelines of Hevner et al. (2004) into account.

2.2.2 Treatment design and treatment validation

The final two tasks of the design cycle are concerned with designing and validating a treatment (artifact). Since this study focuses on developing and validating such an artifact (a maturity model), a supporting research design is selected that defines how to develop such a maturity model theoretically. Mettler (2011) noted that developers of maturity models were not sufficiently supported with existing maturity model design methodologies when making essential decisions on how to design these models.

To tackle these problems, Mettler (2011) developed the so-called 'development cycle'. This development cycle, visualized in Figure 2, consists of the following four phases: define scope, design model, evaluate design, and reflect evolution. Once a need/new opportunity has been identified, corresponding to the problem investigation, the development of a new maturity model starts. All four phases are supported by relevant decision parameters that help the developers of the maturity model with tacking crucial design decisions (Table 2).

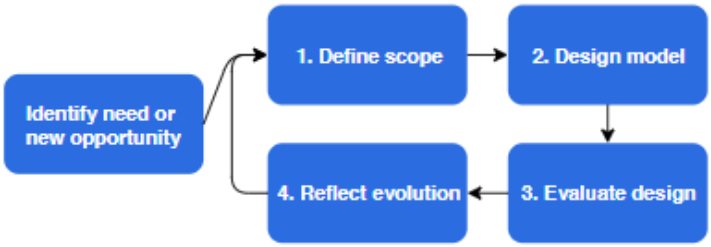


Figure 2 - Development cycle (Adapted from Mettler, 2011)

Phase	Decision parameter	Characteristic			
Define scope	Focus/breadth	General issue		Specific issue	
	Level of analysis/depth	Group decision-making	Organizational considerations	Inter-org. considerations	Global and societal considerations
	Novelty	Emerging	Pacing	Disruptive	Mature
	Audience	Management-oriented	Technology-oriented	Both	
	Dissemination	Open		Exclusive	
Design model	Maturity definition	Process focused	Object focused	People focused	Combination
	Goal function	One-dimensional		Multi-dimensional	
	Design process	Theory-driven	Practitioner-based	Combination	
	Design product	Textual description of form	Textual description of form and functioning	Instantiation (assessment tool)	
	Application method	Self-assessment	Third-party assisted	Certified professionals	
	Respondents	Management	Staff	Business partners	Certified professionals
Evaluate design	Subject of evaluation	Design process	Design product	Both	
	Time-frame	Ex-ante	Ex-post	Both	
	Evaluation method	Naturalistic		Artificial	
Reflect evolution	Subject of change	None	Form	Functioning	Form and functioning
	Frequency	Non-recurring		Continuous	
	Structure of change	External/open		Internal/exclusive	

Table 2 - Decision parameters of the development cycle (Adapted from Mettler, 2011)

Phase 1 and 2 – Define scope & design model (treatment design)

The first two phases of the development cycle, 'define scope' and 'design model', corresponds to the treatment design task of the design cycle and takes the 'design as an artifact' and 'research rigor' guidelines of Hevner et al. (2004) into account. In the first phase of the development cycle, developers of maturity models face the most critical design decisions that influence all remaining decision parameters (Mettler, 2011). Based on the information and knowledge gained with the surveys and the literature review, defining the scope of the maturity model will be done by deciding and elaborating on the following five decision parameters: focus/breadth, level of analysis/depth, novelty, audience, and dissemination.

Once the scope is defined, two activities will be performed to design and develop the maturity model for the DQM domain. First, decisions will be made to the six decision parameters that belong to the design model task of the development cycle: maturity definition, goal function, design process, design product, application method, and respondents. Second, an additional literature study will be performed to study the identified and merged CSFs for DQM from the literature review in detail and populate the maturity model. Additionally, questions, measurement criteria, and maturity levels will be defined, and instantiation in the form of an assessment instrument will be created to make the model fit for the next phase.

Phase 3 – Evaluate design (treatment validation)

The third phase of the development cycle, evaluate design, is "concerned with the verification and validation of the designed maturity model" (Mettler, 2011). This phase corresponds to the treatment validation task of the design cycle since the goal of this task to justify that the designed artifact would contribute to the stakeholder' goals when implemented in the problem context (Wieringa, 2014) and takes the 'design evaluation' and 'research rigor' guidelines of Hevner et al. (2004) into account.

For the validation of the designed maturity model, the assessment instrument of the maturity model is reviewed for content and technical difficulties during a pretest first. Afterward, a case study will be conducted through a maturity assessment at Achmea. The case study protocol of Yin (2003) and the template of Brereton et al. (2008), which is mostly based on the research done by Yin (2003), will be partly applied to shape the case study. The results of this maturity assessment will serve as advice for Achmea but also indicate whether the designed maturity model treats the design problem. The treatment will be considered as fulfilled when the goals of the stakeholders of this study (Ari Sadik and Sven Huijsmans, both supervisors of this thesis project) are satisfied. In other words, this is the case when the designed artifact fulfills the design problem of this study, expressed with the schema defined by Wieringa (2014):

Improve and assess data quality management in an organization **by** developing a maturity model **that** is able to measure the data quality management maturity of a business chain **in order to** get an impression of data quality management maturity.

Phase 4 – Reflect evolution

If the design problem of this study is fulfilled, the design mutability of the maturity model will be contemplated. Although very important, according to Mettler (2011), this final phase is often neglected. The maturity model needs to be refaced from time to time, as the phenomenon being studied is growing (Mettler, 2011). Based on the designing maturity model and the wishes of the stakeholders, the design mutability will be contemplated by deciding and elaborating on the final three parameters: subject of change, frequency, and structure of change. This 'reflect evolution' phase takes the 'communication of research' guideline of Hevner et al. (2004) into account.

2.2.3 Conceptual framework

The conceptual framework, which can be seen in Figure 3 below, gives an overview of the milestones for the critical phases of this thesis project. It also shows the outline of this study by visualizing the coherence between the research questions, research methods, and the deliverables of this study.

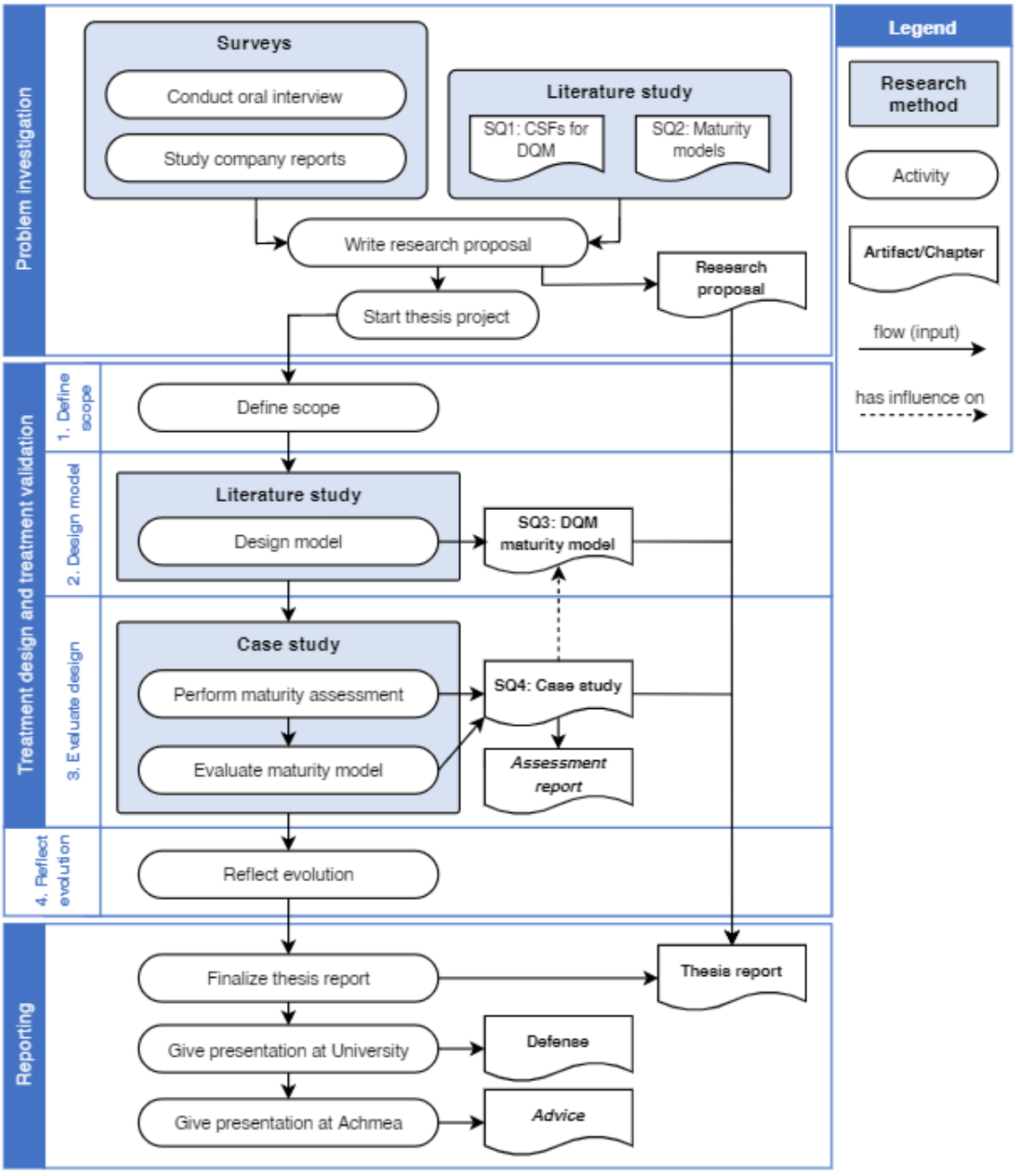


Figure 3 - Conceptual framework

2.3 Literature research protocol

This section discusses the protocol used to find and select relevant literature for this study. A semi-structured approach is defined to ensure that the literature is of high quality and adds value to the goals of this study. Three elements defined by Kitchenham (2014) are used to define the protocol (search strategy, selection criteria, and keywords) and discussed in detail below.

2.3.1 Search strategy

The primary source for collecting relevant literature through a manual search is Google Search. In addition to manual search, the snowballing method of Webster & Watson (2002) is used to structure the process of finding additional related literature. This method requires a starting set of papers from journals and conferences and defines the following protocol:

1. Go backward by reviewing the reference lists of the relevant articles found in the manual search (iterate until no new papers are identified);
2. Go forward by identifying articles citing the articles identified in the manual search and step 1 (iterate until no new papers are identified).

Next to Google Scholar, other search engines are addressed to ensure that all relevant literature is collected. The following three search engines related to Computing and Information Sciences (ACM, DBLP computer science bibliography, and IEEE Computer Society Digital Library) and two general search engines for scientific publications (ResearchGate and SpringerLink) are selected:

- *ACM Digital Library*: Research, discovery, and networking platform which contains scientific publications related to computing and information technology.
- *DBLP computer science bibliography*: Open bibliographic information on computer science proceedings and journals.
- *IEEE Computer Society Digital Library*: Provides access to more than 700 thousand articles on advanced computing.
- *ResearchGate*: Networking site for researchers/scientists to share papers, find collaborators, and ask/answer questions.
- *SpringerLink*: Provides access to scientific publications from books, journals, protocols, reference works, proceedings.

2.3.2 Selection criteria

Selection criteria help to identify primary studies that help to provide evidence about the define research questions and should be decided before conducting the literature to reduce the likelihood of bias (Kitchenham, 2004). The inclusion and exclusion criteria used to identify those primary studies can be found below (Table 3).

Included	Excluded
Studies that are published in English	Studies that have an other publication language than English
Studies related to the concept of DQM or maturity models	Studies with less than 3 pages
Studies that include critical success factors for DQM	Grey literature (non-peer reviewed studies)
Studies that include a maturity model for DQM	Studies that do not answer the research questions
Studies that include a case study (financial/insurance sector)	

Table 3 - Inclusion and exclusion criteria

Despite the changing role of IT and the use of data in organizations due to emerging technologies, no boundary is set regarding the date of publication. To explore the concepts of DQM, CSFs for DQM, maturity models, and DQM maturity models in detail, all relevant literature regarding these topics needs to be studied.

2.3.3 Keywords

To find relevant literature based on the defined search strategy and selection criteria, search strings are formulated to structure the search process. According to Kitchenham (2004), sophisticated search strings can be constructed using the Boolean operators 'AND' and 'OR'. The following search strings were defined and used to find relevant literature in the selected search engines:

- "data quality management"
- "data quality" AND "management" AND (("maturity" OR "capability") AND "model")
- "data quality" AND "management" AND ("critical" AND "success" AND "factors")
- "maturity model" AND ("assessment" OR "situational")

3. Literature Review: Background & Related Work

3.1 Data quality management & critical success factors

SQ1: *What are the critical success factors for data quality management?*

3.1.1 Data quality & data quality dimensions

In scientific literature, there is a consensus that data can be defined as "a set of discrete, objective facts about events" (Davenport & Prusak, 1998). However, there is no consensus on a hard and widely-accepted definition for the concept of 'data quality' (DQ) since what constitutes quality depends on what is considered as good (English, 1999). A selection of DQ definitions cited below show that DQ is a multi-dimensional concept that is often defined as 'fitness for use' or the extent to which the data serves the purpose of the data consumers:

- "We define high-quality data as data that are fit for use by data consumers." (Strong, Lee, & Wang, 1997);
- "The term DQ can best be defined as 'fitness for use'." (Tayi & Ballou, 1998);
- "DQ is defined with two consentient aspects: first, the dependence of perceived quality on the user's needs; second, the so-called 'fitness for use', which is the ability to satisfy the requirements of intended use in a specific situation" (Otto et al., 2007).

The fact that DQ is a multi-dimensional concept is also reflected in the way how the 'fitness for use' of data can be perceived: with DQ dimensions. These dimensions are commonly used to indicate "whether data contains deficiencies or whether there might be DQ issues" (Haegemans, 2018). Over time, a large number of DQ dimensions have been defined. Jayawardene, Sadiq, & Indulska (2015) conducted a systematic review of related research and practitioner literature and identified 127 different DQ dimensions. To create a shared understanding and interpretations of these DQ dimensions, the authors clustered these dimensions into the following eight main clusters (dimensions): completeness, availability & accessibility, currency, accuracy, validity, reliability & credibility, consistency, and usability & interpretability. Table 4 presents the eight main DQ dimensions (supported by a set of dominant quality characteristics, a description, the level of granularity, and the type of the characteristic).

According to Jayawardene, Sadiq, & Indulska (2015), it is essential to consider at which data granularity level the DQ dimensions are applicable. The three granularity levels used by Jayawardene, Sadiq, & Indulska (2015) to classify the quality characteristics are:

1. *Data elements* [E]: an attribute of a real-world entity;
2. *Data record* [R]: a collection of attributes that represent a real-world entity in a database;
3. *Information object* [IO]: a collection of records to accomplish a task.

The quality characteristics are also classified based on the characteristic type, either be one of the following two perspectives (Jayawardene, Sadiq, & Indulska, 2015):

1. *Declarative* [D] perspective: "focuses on user-independent characteristics of data which explains the data itself";
2. *Usage* [U] perspective: focuses on user-dependent characteristics of data related to effective and efficient data creation and usability that contribute to user's judgment about the data's fitness for use".

Dimension	Quality characteristic	Description	Granu-larity	Type
Completeness	Completeness of mandatory values	The attributes which are necessary for a complete representation of a real world entity must contain values and cannot be null.	E	D
	Completeness of optional values	Non-mandatory attributes should not contain invalid null values.	E	D
	Completeness of records	Every real world entity instance that is relevant for the organization can be found in the data.	R	U
	Data volume	The volume of data is neither deficient nor overwhelming to perform an intended task.	IO	U
Availability & accessibility	Continuity of data access	The technology infrastructure should not prohibit the speed and continuity of access to the data for the users.	IO	U
	Data maintainability	Data should be accessible to perform necessary updates and maintenance operations in its entire lifecycle.	R	U
	Data awareness	The data users should be aware of all available data and its location.	IO	U
	Ease of data access	The data should be easily accessible in a form that is suitable for its intended use.	IO	U
	Data punctuality	Data should be available at the time of its intended use.	IO	U
	Data access control	The access to the data should be controlled to ensure it is secure against damage or unauthorized access.	IO	U
Currency	Data timeliness	Data which refers to time should be available for use within an acceptable time relative to its time of creation.	R	U
	Data freshness	Data which is subjected to changes over the time should be fresh and up-to-date with respect to its indented use.	R	U
Accuracy	Accuracy to reference source	Data should agree with an identified source.	E	U
	Accuracy to reality	Data should truly reflect the real world.	R	U
	Precision	Attributes values should be accurate as per linguistics and granularity.	E	D
Validity	Business rules compliance	Calculations on data must comply with business rules.	E	D
	Meta-data compliance	Data should comply with its metadata.	E	D
	Standards and regulatory compliance	All data processing activities should comply with the policies, procedures, standards, industry benchmark practices, and all regulatory requirements that the organization is bound by.	IO	U
	Statistical validity	Computed data must be statistically valid.	IO	U
Reliability & credibility	Source quality	Data used is from trusted and credible sources.	IO	U
	Objectivity	Data are unbiased and impartial.	IO	U
	Traceability	The lineage of the data is verifiable.	R	U
Consistency	Uniqueness	The data is uniquely identifiable.	R	D
	Redundancy	The data is recorded in exactly one place.	R	D
	Semantic consistency	Data is semantically consistent.	E	D
	Value consistency	Data values are consistent and do not provide conflicting or heterogeneous instances.	E	D
	Format consistency	Data formats are consistently used.	E	D
	Referential integrity	Data relationships are represented through referential integrity rules.	R	D

Usability & interpretability	Usefulness and relevance	The data is useful and relevant for the task at hand.	IO	U
	Understandability	The data is understandable.	IO	U
	Appropriate presentation	The data presentation is aligned with its use.	IO	U
	Interpretability	Data should be interpretable.	IO	U
	Information value	The value that is delivered by quality information should be effectively evaluated and continuously monitored in the organizational context.	IO	U

Table 4 - Data quality dimensions and quality characteristics (Adapted from Jayawardene, Sadiq, & Indulska (2015))

3.1.2 Data quality management

As with the DQ concept, there is also no consensus on a hard and widely-accepted definition in scientific literature for the 'data quality management' (DQM) concept. To identify common aspects in DQM definitions, a selection identified in the literature is cited below:

- "We refer to DQM as quality-oriented DM, i.e., DM focusing on the collection, organization, storage, processing, and presentation of high-quality data" (Otto et al., 2007);
- "DQM focuses on the planning, provisioning, organization, usage, and disposal of high-quality data" (Weber et al., 2009);
- "We define DQM as a set of coordinated activities to direct and control an organization with regard to DQ" (Lucas, 2010);
- "DQM as an organizational function comprises all practices, methods, and systems for analyzing, improving, and maintaining the quality of data. DQM basically aims at maximizing the value of data." (Ofner, Otto, & Österle, 2013);
- "DQM consists of applying the concept and practices of Total Quality Management (TQM) to improve the quality of data and information, involving the definition of policies and rules, assessment of DQ (including auditing and certification), data analysis, cleaning, and correction, DQ improvement and education" (Santos & Lucas, 2019).

The definitions show that managing DQ is not just only focusing on improving DQ by performing ad-hoc repair activities: it involves applying multiple concepts, practices, and activities to become successful in DQM. This view is also reflected by Lucas (2019), which states that effective DQM "should transcend the activities of fixing non-quality data to prevent DQ problems by managing over its life cycle". This helps an organization to become effective in meeting the information needs of their stakeholders. Cultural change, control and allocation of resources, demanding leadership, and authority is required to create effective collaboration between business and IT units to address both organizational and technical perspectives (Lucas, 2019).

These requirements for effective DQM assume that data governance (DG) should be applied within an organization. Mosley (2008) defined DG as "the exercise of authority, control and shared decision making (planning, monitoring, and enforcement) over the management of data assets" (Mosley, 2008). However, since DG does not equal DQM in two aspects (concerning the entity that makes the decisions, or to their scope), DG is seen as a discipline within the DQM domain. The concept of DG itself is more focused on the corporate environment and executive management, which includes areas beyond the DQM domain (e.g., data security, privacy, information life-cycle management) (Lucas, 2019). Also, DG is the responsibility of directors and executive management, while DQM is the responsibility of lower-level management and executive staff (Lucas, 2019).

3.1.3 Data quality management critical success factors: identification, ordering, and merging

Critical success factors for data quality management identified and ordered by Santos & Lucas (2019)

Santos & Lucas (2019) conducted an extensive literature review in which they identified 24 so-called Critical Success Factors (CSFs) for DQM. These CSFs are essential elements for an organization or project to achieve its mission and should be applied within an organization (Huang, Wu, & Chen, 2013) to ensure a competitive and successful performance (Rockart, 1979).

The identified CSFs were ordered in a Delphi study by a panel of experts (specialists in the area of knowledge). In a Delphi study, "each questionnaire should correspond to a round, making as many rounds as necessary to obtain a consensus or the confirmation that the consensus is not possible" (Santos & Lucas, 2019). During the first round of the Delphi study, the experts added 'data governance' to the list of CSFs. This factor is defined as "the set of essential actions to ensure data compliance with organizational strategies" (Santos & Lucas, 2019). After the second round, a consensus was reached to a final order of the 25 CSFs (Table 5). The CSFs are provided with a code for reference purposes, based on the authors' initials and the order in the final list.

CSFs for DQM by Santos & Lucas (2019)			
[SL-1]	DQ policies and standards	[SL-14]	Appointment of managers and definition of roles
[SL-2]	Input controls	[SL-15]	Documentation
[SL-3]	Production of a strategic plan for DQ	[SL-16]	Communication
[SL-4]	Organizational culture with a focus on DQ	[SL-17]	Middle management commitment and support
[SL-5]	Top management commitment and support	[SL-18]	Teamwork
[SL-6]	Data governance	[SL-19]	Security and internal control
[SL-7]	Continuous improvement	[SL-20]	Risk Management
[SL-8]	Internal and external monitoring and evaluation	[SL-21]	Sufficient resources
[SL-9]	Change Management	[SL-22]	Storage Management
[SL-10]	Conducting regular audits	[SL-23]	Evaluate cost/benefit tradeoffs
[SL-11]	Architecture Management	*[SL-24]*	*Effective relationship with employees*
[SL-12]	User Focus	*[SL-25]*	*Physical environment*
[SL-13]	Education and Training		

Table 5 - CSFs for DQM identified by Santos & Lucas (2019)

According to the study of Santos & Lucas (2019), the most important CSF for DQM is 'DQ policies and standards'. It entails the standardization of codes, rules, and definitions, reformulation of the data model (only if necessary), and the implementation of a standard methodology (Santos & Lucas, 2019). The first five CSFs maintained their position in the two evaluation rounds of the Delphi study: only the order of these CSFs was changed. The final two CSFs, 'effective relationship with employees' and 'workplace environment', maintained their 24th and 25th position during the evaluation rounds. According to Santos & Lucas (2019), this seems that these CSFs are not considered relevant to the DQM domain.

Santos & Lucas (2019) also explicitly highlighted the importance of 'data governance'. This CSF was not identified in the literature study but was added to the list after the first evaluation round. In the end, it is ordered in the 6th position, which demonstrates the convergence of opinions regarding its importance. This CSF indeed shows that DG is interrelated with DQM but seen as a

separate discipline within the DQM domain, as discussed in the previous subsection. More information of the CSFs for corporate DQM can be found in Appendix II-A.

Critical success factors for corporate data quality management identified and ordered by Lucas (2019)

Besides the study of Santos & Lucas (2019), there is a second recent study published in which 22 CSFs for corporate DQM are identified (Lucas, 2019). In a focus group, a consensus among the academic and practitioners was reached that four CSFs ('physical environment', 'storage management', 'employee relations', and 'nature of the IS') should not be considered as CSFs but as contingency factors. These contingency factors are variables "that moderates the effect of an organizational characteristic on organizational performance" (Donaldson, 2001).

Just like in the study of Lucas & Santos (2019), it was decided to add 'data governance' to the initial list of CSFs. The participants also reached a consensus to merge ('organization for quality' and 'teamwork' is included in 'data governance', and 'strategic data quality planning' is included in 'management commitment and leadership') and rename some CSFs which resulted in a final list of 19 CSFs for corporate DQM. After the focus group and the execution of the Delphi study, a consensus was reached for a final order of the 19 CSFs (Table 6). The CSFs are provided with a code for reference purposes, based on the authors' initials and the order in the final list.

CSFs for corporate DQM by Lucas (2019)			
[L-1]	Data Governance (+ 'organization for quality' and 'teamwork')	[L-11]	Personnel Competency
[L-2]	Management Commitment and Leadership (+ 'strategic data quality planning')	[L-12]	Management of Changes
[L-3]	Continuous Data Quality Management Improvement (+ 'continuous improvement' and 'DQ controls/input controls')	[L-13]	Data Security Management
[L-4]	Data Architecture Management	[L-14]	Data Quality Risk Management
[L-5]	Culture and Communication	[L-15]	Audit and Reviews
[L-6]	Data Quality Policies and Standards	[L-16]	Training
[L-7]	Data Quality Assessment/Monitoring	[L-17]	Understanding of the IS and the relevance of DQ
[L-8]	Data Quality Requirements Management	[L-18]	Evaluate Cost/Benefit Trade-offs
[L-9]	Focus on Data Customer Satisfaction	[L-19]	Supplier Partnership
[L-10]	Data Product Lifecycle Management		

Table 6 - CSFs for corporate DQM identified by Lucas (2019)

According to this study, the most important CSF for corporate DQM is 'data governance'. This study defines DG as "a set of key actions to ensure data compliance with organizational strategies" (Lucas, 2019). Despite the use of 'key actions' instead of 'essential actions' used in the study of Santos & Lucas (2019), both studies use the same definition.

However, in comparison to the paper of Santos & Lucas (2019) in which only the definition of DG is given, the paper of Lucas (2019) provides additional information to this CSF: "It defines a suitable organizational structure to produce high-quality information. It should define responsibilities for the DQ: identify the owners and custodians; appoint data stewards and a data champion; appoint an expert or a group of experts as DQ managers. It should promote teamwork between business and IT people, as a key to improving data quality". More information of the CSFs for corporate DQM can be found in Appendix II-B.

Merging the critical success factors identified by Santos & Lucas (2019) and Lucas (2019)

As shown in Tables 5 and 6, it appears that multiple CSFs with similar intent are in both lists. To address this issue, CSFs that have similar intents have been merged based on the information of section 3.1.3, information in the papers of Lucas (2019) and Santos & Lucas (2019), and the knowledge gained during the literature study. The results of this merging process are outlined in Table 7. The last column of this table defines why certain CSFs are (not) merged:

- *Same descriptors (SD)*: the descriptor of a CSF identified by Santos & Lucas (2019) is exactly or almost the same as the descriptor of a CSF identified by Lucas (2019), or vice versa;
- *Different descriptors but similar intent (SI)*: the descriptor of a CSF identified by Santos & Lucas (2019) is not the same as identified by Lucas (2019), or vice versa, but both CSFs appeared to have similar intent;
- *Consistent with consensus (CWC)*: in the study of Lucas (2019), experts (recognized academics and practitioners) reached consensus to include certain CSFs in a different CSF. To keep the domain components for the DQM3 consistent with the consensus reached by these experts, CSFs identified by Santos & Lucas (2019) that correspond to the included CSFs in the study of Lucas (2019) are merged;
- *No similarity with other CSFs (X)*: the CSF is not merged with other CSFs because there are no CSFs with similar intent. CSFs with no similarity are either kept the same or written slightly differently to increase consistently with other CSFs.

CSFs for DQM		Merged / renamed CSF	Why merged?
[SL-1] [L-6]	DQ policies and standards Data Quality Policies and standards	Data quality policies and standards	SD
[SL-20] [L-14]	Risk Management Data Quality Risk Management	Data quality risk management	SD
[SL-11] [L-4]	Architecture Management Data Architecture Management	Data architecture management	SD
[SL-23] [L-18]	Evaluate cost/benefit tradeoffs Evaluate Cost/Benefit Trade-offs	Evaluate cost/benefit tradeoffs	SD
[SL-12] [L-9]	User Focus Focus on Data Customer Satisfaction	Data customer satisfaction	SI
[SL-5] [SL-18] [SL-14] [L-2]	Data governance Teamwork Appointment of managers and definition of roles Data Governance (+ 'organization for quality' and 'teamwork')	Data governance	SD, CWC
[L-5] [SL-4] [SL-16]	Culture and Communication Organizational culture with a focus on DQ Communication	Culture and communication	SD, SI
[L-11]	Personnel competency	Personnel competency	X
[L-8]	Data Quality Requirements Management	Data quality requirement management	X
[L-19]	Supplier Partnership	Data supplier quality management	X

[SL-5] [SL-17] [SL-21] [SL-3] [L-17] [L-2]	Top management commitment and support Middle management commitment and support Sufficient resources Production of a strategic plan for DQ Understanding of the IS and the relevance of DQ Management Commitment and Leadership (+ 'strategic data quality planning')	Management commitment, leadership, and support	SD, SI, CWC
[SL-9] [L-12]	Change Management Management of Changes	Change management	SI
[SL-15]	Documentation	Documentation	X
[L-16] [SL-13]	Training Education and Training	Education and training	SD
[L-10]	Data Product Lifecycle Management	Data product lifecycle management	X
[SL-19] [L-13]	Security and internal control Data Security Management	Data security management	SI
[L-3] [SL-7] [SL-2]	Continuous Data Quality Management Improvement (+ 'continuous improvement' and 'DQ controls/input controls') Continuous improvement Input controls	Continuous data quality (management) improvement	CWC
[L-7] [SL-8] [SL-10] [L-15]	Data Quality Assessment/Monitoring Internal and external monitoring and evaluation Conducting regular audits Audit and Reviews	Data quality assessment/monitoring	SI

Table 7 - Merged CSFs for DQM

Result of merging the critical success factors for data quality management (answer on SQ1)

In the study of Lucas (2019), the focus group reached consensus that 'storage management' should not be considered as a CSF but as a contingency factor. To be consistent with the consensus reached, contingency factors are excluded as CSFs. This results in the following list of merged CSFs (Table 8), provided with a code for future references (in alphabetical order):

Critical success factors for DQM			
CM	Change management	DQREM	Data quality requirements management
CDQI	Continuous data quality (management) improvement	DQRIM	Data quality risk management
CC	Culture and communication	DSM	Data security management
DAM	Data architecture management	DSQM	Data supplier quality management
DCS	Data customer satisfaction	D	Documentation
DG	Data governance	ET	Education and training
DPLM	Data product lifecycle management	ECBT	Evaluate cost/benefit tradeoffs
DQAM	Data quality assessment/monitoring	MCLS	Management commitment, leadership, and support
DQPS	Data quality policies and standards	PC	Personnel competency

Table 8 - Critical success factors for data quality management

3.1.4 Data quality management critical success factors: other studies

Although the first sub question has already been answered in the previous subsection, two additional studies are identified that have investigated the impact of a variable on the importance and performance of DQ in account information systems and a study grouping CSFs for DQM and a study grouping CSFs for DQM. These three studies are discussed in detail to get a complete picture of existing scientific literature related to CSFs for DQM.

Influence of an organizations' size and industry on CSFs for DQM

Two studies have investigated the impact of a variable on the importance and performance of DQ in accounting information systems. The study of Xu (2003) attempted to reveal whether organizations of different sizes consider these factors different. The study of Xu & Lu (2003) focused on revealing differences between different industries.

In both studies, a national wide survey was conducted in Australia. In the study of Xu (2003), organizations were categorized based on the annual revenue of the organization in the following categories: very small organizations (< $5 million); small organizations ($5 million to $9 million); medium-sized organizations ($10 million to $99 million); and large organizations (> $100 million). The answers of the survey were analyzed, and the results showed that only "significant differences were found in regarding the importance of 'internal controls' and the performance of 'audit and reviews' between the organizations that had different revenues" (Xu, 2003). A Tukey Post Hoc analysis showed that only significant differences were found between the subgroups in the performance of the factor 'audit and review'. Tukey tests reveal that only significant differences were found in the performance of 'audit and review' between very small organizations, and small and medium-sized organizations.

In the study of Xu & Lu (2003), the respondents were asked to select the industry in which they were working. The manufacturing industry was most represented (34.4%) compared to services (27.9%) and the finance & insurance industry (6.6%). The factors that were found to have significant differences between industry groups are: 'data supplier quality management', 'continuous improvement', and 'risk management' have differences regarding the importance; 'user focus' and 'employee relations' have differences regarding the performance. After further analysis, no significant effect was found which distinct the finance & insurance industry compared to other industries. However, the results of the survey indicate that the surveyed organizations were aware of the importance of the CSFs that have an impact on DQ of accounting information systems but "comparing to their consideration of the importance of the factors, actual performance of these factors is not up to the satisfactory level" (Xu & Lu, 2003).

Categorization of data quality management critical success factors

The study of Xu (2013) is the only study that tried to combine CSFs for DQM into categories. For this study, Xu developed a CSF model of information systems' DQ in which factor categories are defined that potentially can influence DQ in information systems: information system (IS) characteristics, DQ characteristics (factors that are directly related to DQ), stakeholders' related factors (people/human factors), organizational factors, and external factors (factors over which the organization has no control).

Although the model was built based on scientific literature and previous case studies, Xu (2013) conducted a large-scale survey to further develop and test the model. To identify factor groups from the 25 CSFs identified from earlier research, the results of the survey were used to perform a factor analysis. The factor analysis produced four factorial groups (DQM factors, people factors, organizational factors, and environmental factors) and re-grouped the 25 CSFs identified earlier as follows:

- *DQM factors*: top management commitment, middle management, education & training, DQ vision, DQ control, input control, user focus, nature of the IS, change management;
- *People & assessment factors*: employee factors, measurement report, data supplier quality management, continuous improvement, teamwork, evaluate cost/benefit tradeoffs, understanding of the systems and DQ, risk management;
- *Organizational factors*: role of DQ manager, organizational structure, policies & standards, organizational culture;
- *Environmental & personnel factors*: personnel competency, physical environment.

Main takeaway of the other studies on critical success factors discussed above

In both the publications of Xu (2003) and Xu & Lu (2003), no significant effects were found concerning to the effect of the industry and the size of an organization on the importance of the CSFs (Xu & Lu, 2003). The main message that can be drawn from these results is that it should not be a problem to apply a DQM maturity model (populated with the merged CSFs) to different industries or organizations of various sizes, without taking into account the degree of importance of the factors.

3.2 Maturity models & data quality management

SQ2: W*hich of the existing maturity models allows an organization to perform a self-assessment in which the data quality management maturity level of a business chain is measured?*

3.2.1 Maturity model

According to Proença & Borbinha (2016), a maturity model "is a technique that has been proved to be valuable in measuring different aspects of a process or an organization". Becker, Knackstedt, & Pöppelbuß(2009) define the concept of the maturity model as an artifact that aims to solve the problem of defining an organization concerning their current status of capabilities and deriving means for improvement (Spruit & Pietzka, 2015). When applying such a model, an organization is provided with (Proença & Borbinha, 2016):

1. A measuring for auditing and benchmarking;
2. A measuring of progress assessment against objectives;
3. An understanding of strengths, weaknesses, and opportunities. This understanding can support decision making in the fields of project portfolio management and strategy.

Despite the initial application of the 'maturity model' concept in the 1970s, the popularity of these models has increased within the field of information systems since the introduction of the Capability Maturity Model (CMM) (Mettler & Rohner, 2009). This CMM was developed to present sets of practices in several key software process areas (Paulk et al., 1993). It defines five maturity levels (initial, repeatable, defined, managed, and optimizing) and is recognized as the standard in the world of maturity models. The CMM later evolved to Capability Maturity Model Integration (CMMI), which defines five maturity levels (performed, managed, defined, quantitatively managed, and optimizing).

With the intensification of maturity models, different types of maturity models are proposed over time. However, according to Fraser, Moultrie, & Gregory (2002), most maturity models "share the common property of defining a number of dimensions or process areas at several discrete stages or levels of maturity, with a description of characteristic performance at various levels of granularity." Fraser, Moultrie & Gregory (2002) identified six common elements in maturity models:

1. Number of levels or stages (typically three to six);
2. Descriptor for each level (in CMM: initial, repeatable, defined, managed, and optimizing);
3. Generic description or summary of the characteristics of each level as a whole;
4. Number of dimensions or 'process areas';
5. Number of elements or activities for each process area; and
6. Description of each activity as it might be performed at each maturity level.

3.2.2 Maturity model comparison framework

Mettler et al. (2010) developed a classification framework for information system maturity models. This framework can be applied to analyze and compare existing maturity models. It consists of the following three dimensions: (1) 'general model attributes' that describe basic characteristics, (2) 'maturity model design' attributes that are related to the construction and organization of the model, and (3) 'maturity model use' attributes that describe the usage of the maturity model. All three dimensions are discussed in detail below.

Dimension I – General model attributes

The general model attributes are easy to define without the need to have a deep knowledge of the content of the maturity model and are used to rapidly give potential users of the maturity model an overall overview (Mettler et al., 2010). The attributes for this dimension are: name, acronym, primary source, secondary source, addressed topic, origin, audience, year of publication, and access. However, Mettler et al. (2010) added remarks on some of these general model attributes:

1. The acronym and secondary sources should only be added if it exists;
2. The addressed topic should be categorized in terms of the IS framework of Bacon & Fitzgerald (2001);
3. The origin of the topic is either 'academic' or 'practice'.
4. The audience is either 'management-oriented', 'technology-focused' or 'no clear distinction'.
5. The mode of access is either 'free' or 'paid'.

Dimension II – Maturity model design

The following four attributes in the 'maturity model design' dimension can be used to describe the form and organization of a maturity model (Mettler et al., 2010): concept of maturity, composition, reliability, and mutability. These four attributes are discussed in detail below.

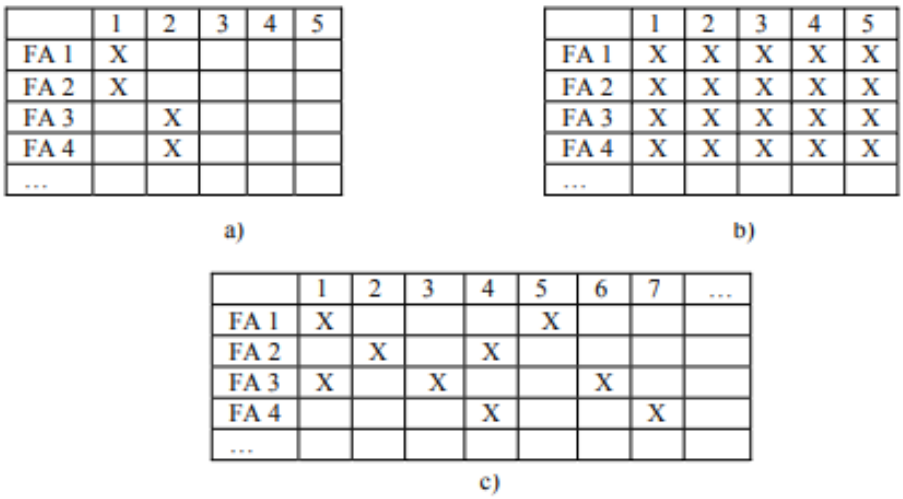
Concept of maturity The 'concept of maturity' attribute is in most instances reflected on a one-dimensional manner, either focusing on: 'process maturity' (the extent to which a process is explicitly defined, managed, measured, controlled, and effective (Chrissis, Konrad, & Shrum, 2003)), (2) 'object maturity' (the extent to which an object reaches a predefined level of sophistication (Gericke, Rohner & Winter, 2006)), or 'people capability' (the extent to which the workforce within an organization can enable knowledge creation and enhance proficiency (Gillies & Howard, 2003)).

Composition part A

Based on the common maturity model elements described earlier, Fraser, Moultrie & Gregory (2002) made a distinction between three types of models which are used as conditions in the classification system of Mettler et al. (2010): (1) 'maturity grids' (only contains text descriptions for each activity at each maturity level, mostly a few pages of text); (2) 'CMM-like models' (models that repeat the structure of CMM, often entail a greater complexity then maturity grids due to wide range of scales and subscales); or (3) 'Likert-like questionnaires' (evaluate relative performance on a scale from 1 to n, equivalent to a maturity grid in which only the characteristics of the top level are described).

Composition part B

Van Steenbergen et al. (2007) also identified two different types of maturity models that repeat the structure of CMM-like models: 'staged model', and 'continuous model'. These models mostly distinguishes between five maturity levels. However, in the staged maturity model, several focus areas are defined which are specific for a certain maturity level. All focus areas that are defined to that specific level have to be implemented for the organization to achieve that maturity level. The continuous model is different from the staged model in the fact that the focus areas are not attributed to a maturity level, but within each focus area, the five levels are distinguished. Also, a third type of model is identified that departs from the idea that there are five generic maturity levels: (c) 'focus area oriented model'. In this type of model, every focus area has its number of specific maturity levels. Figure 4 illustrates the differences between the types of models identified by Van Steenbergen et al. (2007).



	1	2	3	4	5
FA 1	X				
FA 2	X				
FA 3		X			
FA 4		X			
...					

a)

	1	2	3	4	5
FA 1	X	X	X	X	X
FA 2	X	X	X	X	X
FA 3	X	X	X	X	X
FA 4	X	X	X	X	X
...					

b)

	1	2	3	4	5	6	7	...
FA 1	X				X			
FA 2		X		X				
FA 3	X		X			X		
FA 4				X			X	
...								

c)

Figure 4 - Types of maturity models (Reprinted from Van Steenbergen et al., 2007)

Reliability

According to Mettler et al. (2010), reliability is important to enhance the reusability of a maturity model. Mettler et al. (2010) distinguishes in this attribute between two conditions: 'verification' ("the process of determining that" a maturity model "represents the developer's conceptual description and specifications with sufficient accuracy" (Conwell, Enright, & Stutzman, 2000)), or 'validation' ("the process of determining the degree to which" the maturity model "is an accurate representation of the real world from the perspective of the intended uses of the model" (Conwell, Enright, & Stutzman, 2000)).

Mutability As discussed before, the design mutability of the maturity model must be contemplated: the maturity model has to be refaced periodically because the phenomenon under study is growing (Mettler, 2011). Two conditions are proposed by Mettler et al. (2010) for this attribute: (1) 'form' (underlying meta-model or model schema, descriptions of the maturity levels or question items); or (2) 'functioning' (how maturity is assessed).

Dimension III – Maturity model use

The following three attributes in the 'maturity model use' dimension can be used to describe the usage of a maturity model: method of application, support of application, and practicality of evidence. These attributes are discussed in detail below.

Method of *application* According to Mettler et al. (2010), an important criterion for the selection of a maturity model is the method of how data is collected to assess maturity. Based on De Bruin & Rosemann (2015), Mettler et al. (2010) distinguishes three distinct approaches in assessing maturity: (1) 'self-assessment' (refers to the systematic process of gathering information about the own capabilities and level of maturity); (2) 'third-party assisted' assessment (slight modification of self-assessment, supported by external specialists); and (3) 'certified practitioners' (assessment outsourced to certified practitioners).

Support of of application Mettler et al. (2010) differentiate between three stages of assistance concerning the support of the model application: 'no supporting materials', 'textual description or handbook', or 'software assessment tool'.

Practicality of evidence This attribute distinguishes between two conditions: 'explicit recommendations' (telling exactly what to do), and 'implicit improvement activities' (suggestions for improvements are made).

Template to classify information system maturity models

Based on three dimensions, attributes, defined conditions identified by Mettler et al. (2010) and Van Steenbergen et al. (2007), a template is designed (Table 9). This template will be used to compare the existing DQM maturity models.

Dimension	*Attribute*	*Conditions*		
General model attributes	Name	[name]		
	Acronym	[acronym]		
	Primary source	[primary source]		
	Secondary source	[secondary source]		
	Addressed topic	[addressed topic]		
	Origin	Academic	Practice	
	Audience	Management-oriented	Technology-focused	No clear distinction
	Year of publication	[year of publication]		
	Access	Free	Paid	

Maturity model design	Concept of maturity	Process maturity	Object maturity	People capability
	Composition - A	Maturity grid	CMM-like	Likert-like questionnaire
	Composition – B	Staged model	Continuous model	Focus area model
	Reliability	Validation	Verification	
	Mutability	Form	Functioning	
Maturity model use	Method of application	Self-assessment	Third-party assisted assessment	Certified practitioners
	Support of application	No supporting materials	Textual description or handbook	Software assessment tool
	Practicality of evidence	Implicit improvement activities	Explicit recommendations	

Table 9 - Template to classify a maturity model (Adapted from Mettler et al. (2010) and Van Steenbergen et al. (2007))

3.2.3 Data quality management maturity models: identification

According to Lucas (2010), limited research has been conducted into instruments that assess the progress and performance of DQM initiatives. To the best knowledge of Lucas (2010), only the following three maturity models were developed up to 2010 for the DQM domain:

[1] – "A data quality management maturity model" – Ryu, Park, & Park (2006)

This maturity model reflects the DQM domain in three different viewpoints: total corporate integration point of view, data structure quality management, and maturity stages point of view. This model defines four maturity levels ('initial', 'defined', 'managed', and 'optimized'), explains the issues at these levels, and suggests some solutions to these issues.

[3] – "IQM3: information quality management maturity model" – Caballero et al. (2008)

The 'Information Quality Management Maturity Model' (IQM3) is one of the components of the 'Information Quality Management Framework' (IQMF). The IQM3 is based on the staged maturity levels from the CMMI and defines five maturity levels ('initial', 'defined', 'integrated', 'quantitative managed', and 'optimizing'). Every maturity level defines 'Key Process Areas (KPAs)' which are supported by an information quality management goal.

[4] – "Towards a maturity model for corporate data quality management" – Hüner, Ofner, & Otto (2009)

The maturity model for corporate data quality management (CDQ MM) is developed based on the framework of corporate DQM. In this maturity model, the concept is subdivided into two generic sub-concepts: a 'domain reference model' (which shows the scope or domain that is assessed), and an 'assessment model' (which defines how the maturity levels are assigned to elements of the domain reference model).

Besides the three maturity models introduced above, four additional maturity models are identified after executing the literature review protocol in section 2.2:

[2] – "IQM-CMM: a framework for assessing organization information quality management capability maturity" – Baškarada, Koronios, & Gao (2007)

The Information Quality Management Capability Maturity Model (IQM-CMM) is based on a set of criteria that are thought to be of importance when considering a holistic approach for information quality management. The IQM-CMM defines five maturity levels: 'chaotic', 'reactive', 'measuring', 'managed', and 'optimizing'.

[5] – "The Practitioners Guide to Data Quality Improvement" – Loshin (2011)

In the practitioners guide written by Loshin (2011), a DQ framework is defined which looks at varying degrees of maturity concerning eight components related to DQ: DQ expectations, dimensions of DQ, policies, procedures, governance, standards, technology, and performance management. These eight components from the DQ framework are mapped to the Data Quality Capability/Maturity model and defines five levels of maturity: (1) initial, (2) repeatable, (3) defined, (4) managed, and (5) optimized.

[6] – "A maturity model for enterprise data quality management" – Ofner, Otto, & Österle (2013)

Ofner, Otto, & Österle (2013) presents a maturity model for enterprise DQM. This model uses the European Foundation for Quality Management (EFQM) Excellence model as the base model. This paper also shows a conceptual model. No fixed maturity level has been defined for this model.

[7] – "A maturity model for improving data quality management" – Kirikoglu (2017)

Kirikoglu (2017) presents a maturity model for the DQM domain which is supported by a scorecard. This scorecard can be used to determine the current maturity state of a firm and consists of 12 DQM factors. For the development of the model, the CMM model of Paulk et al. (1993) and the maturity model of Ryu et al. (2006) were used as a starting point. This maturity model defines five maturity levels: 'person dependent & basic', 'policies, standards, and procedures', 'defined and stable', 'managed and standardized', and 'continuous improvement'.

The maturity models introduced above are provided with a number between brackets (e.g., [1], [2], [3], etc.) based on the year of publication. These numbers are added for reference purposes and will correspond to the numbers in Table 10.

3.2.4 Data quality management maturity models: comparison

To compare the existing maturity model identified and introduced in the previous subsection, the template presented in Table 9 is used to classify the models. Since this literature study only focuses on the identification of scientific maturity model developed specifically for the DQM domain, the following attributes of the template are omitted because they all contain the same values: addressed topic (value is 'DQ/DQM'), origin (value is 'academic'), and access (value is 'free'). The results of the classification can be found below in Table 10.

Limitations of existing maturity models for the data quality management domain (answer on SQ2)

Only four maturity models made it possible to determine the used 'method of application': [2] the IQM-CMM model of Baškarada, Koronios, & Gao (2007); [4] the CDQ MM of Hüner, Ofner, & Gao (2007); [6] the maturity model for enterprise DQM of Ofner, Otto, & Österle (2013); and [7] the DQM maturity model of Kirikoglu (2017). The other three models do not give any information about the 'method of application, so these models do not fulfill the requirements of the problem statement: performing a self-assessment.

	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Acronym	X	IQM-CMM	IQM3	CDQ MM	X	X	X
Primary source	Ryu, Park, & Park (2006)	Baškarada, Koronios, & Gao (2007)	Caballero et al. (2008)	Hüner, Ofner, & Otto (2009)	Loshin (2011)	Ofner, Otto, & Österle (2013)	Kirikoglu (2017)
Secondary source	X	X	X	http://cdq.iwi.unisg.ch/	X	X	X
Audience	Technology-focused	Management-oriented	Management-oriented	Management-oriented	No clear distinction	Management-oriented	No clear distinction
Year of publication	2006	2007	2008	2009	2011	2013	2017
Concept of maturity	Object & process	Process & object	Process & object	Process & object	Process & object	Process & object	Process & object
Comp. - A	?	Maturity grid	CMMI-like	CMM-like	Maturity grid	CMM-like	Maturity grid & Likert-like
Comp. - B	Staged	Staged	Staged	Continuous	Staged	Continuous	Continuous
Reliability	Verified	Verified	?	Verified	?	Verified	Verified
Mutability	Form (maturity levels)	Form & functioning (partial)	Form & functioning (partial)	Form & functioning (meta-model)	Form (maturity levels)	Form (conceptual model)	Form & functioning
Method of application	?	Self-assessment	?	Self-assessment	?	Self-assessment	Self-assessment
Support of application	No supporting materials	No supporting materials	No supporting materials	No supporting materials	No supporting materials	No supporting materials	Scorecard
Practicality of evidence	Explicit recommendations	Implicit improvements	Explicit recommendations	Explicit recommendations	Implicit improvements	Implicit improvements	Implicit improvements

Table 10 - Classification of data quality management maturity models

The four maturity models that are classified with the 'self-assessment' condition should provide additional supporting materials that can be used to measure the DQM maturity level. The classification highlights that only one of the four maturity models has a different condition in the 'support of application' attribute than 'no supporting materials': [7] the DQM maturity model of Kirikoglu (2017) which provides a scorecard. This scorecard has several limitations which shows that also this maturity model does not fulfill the requirements of the problem statement:

1. The scorecard only includes a limit amount of 12 factors. Factors such as management commitment and DQ assessments are ignored, while the CSFs identified in section 3.1.3 shows that these are important to become successful in the management of DQ and should not be ignored;
2. The factors are too generic (e.g., "The firm is focused on continuously improving" and "The firm has automated their procedures as much as possible"); and
3. The scorecard is not applicable to measure the maturity level of a business chain because it takes a 'firm' as scope.

Based on the classification and the limitations in the exiting DQM maturity models, it can be concluded that there exist no maturity model that allows an organization to perform a self-assessment in which the DQM maturity level of a business chain is measured. The development of a maturity model based on the identified CSFs for DQM can therefore lead to interesting scientific insights.

4. Define Scope & Design Model (Treatment Design)

SQ3: *How could the identified critical success factors for data quality management be combined in a maturity model to measure the data quality management maturity level of a business chain?*

This chapter presents the proposed solution to the problem statement: the 'Data Quality Management Maturity Model' (DQM3). Section 4.1 is concerned with defining the scope of the maturity model. The final section of this chapter, section 4.2, presents the developed domain reference model and assessment model. These two models together form the DQM3.

4.1 Define scope

For defining the scope of a maturity model, Mettler (2011) defined the following five decision parameters: focus/breadth, level of analysis/depth, novelty, audience, and dissemination. The decisions made for these parameters are discussed in detail below and summarized in Table 11.

Focus The first decision parameter of this phase is characterized as a 'specific issue'. DQM is defined as a "quality-oriented DM, i.e., DM focusing on the collection, organization, storage, processing, and presentation of high-quality data" (Otto et al., 2007).

Level of analysis The 'level of analysis' decision parameter is characterized as both 'group decision-making' and 'organizational considerations' because the model will focus on measuring the DQM maturity level of a business chain.

Novelty Taking into account the current situation in the scientific literature regarding CSFs for DQM (limited attention until recent publications by Santos & Lucas (2019) and Lucas (2019)) and DQM maturity models, the 'novelty' decision parameter is characterized as 'emerging'.

Audience 'Audience' is characterized as both 'management-oriented' and 'technology-focused' because the identified and merged CSFs for DQM in section 3.1.3 addresses both audiences: e.g., '[MCLS] Management commitment, leadership and support' addresses a more management-oriented audience, while '[DAM] Data architecture management' focuses more on a technology-focused audience.

Dissemination The 'dissemination' decision parameter is characterized as 'open', because this thesis report is published in an online database which can be accessed by everyone. This complies with the policy of Utrecht University.

Decision parameter	***Characteristic***			
Focus/breadth	General issue		Specific issue	
Level of analysis/depth	Group decision-making	Organizational considerations	Inter-org. considerations	Global and societal considerations
Novelty	Emerging	Pacing	Disruptive	Mature
Audience	Management-oriented	Technology-focused	Both	
Dissemination	Open		Exclusive	

Table 11 - Define scope decision parameters and characteristics

4.2 Design model

After defining the scope of the DQM3, the actual maturity model can be designed. Mettler (2011) defined the following six decision parameters that help make critical design decisions: maturity definition, goal function, design process, design product, application method, and respondents. The decision made for these parameters are discussed below and summarized in Table 12.

Maturity definition

According to Mettler (2011), it is "important to have a clear understanding of what is meant by 'maturity'. A process-focused definition "implies to center on activities and work practices (e.g., inputs/outputs of specified tasks) to define more effective procedures" (Mettler, 2011). When taking an object-focused definition, "features of work products (e.g., functional and non-functional requirements) are investigated to enhance their mode of operations" (Mettler, 2011). A people-focused definition is used when "emphasis of the model lies more on the soft capabilities (e.g., people's feelings and behavior)" (Mettler, 2011). The DQM3 addresses a 'combination' of these three maturity definitions since the DQM domain also addresses all those definitions.

Goal function

Mettler (2011) highlights that the progress of maturity can be either defined "as 'one-dimensional' (i.e., solely focusing on one target measure like efficiency) or 'multi-dimensional' (i.e., focusing on multiple, sometimes divergent goals or competitive bases)". Just as the 'maturity definition' is applied in a multi-dimensional way, the 'goal function' is also characterized as 'multi-dimensional' since the CSFs for DQM are too different in nature to reflect them in just one dimension.

Design process

According to Mettler (2011), the nature of the design process has to be determined "to identify the knowledge base for deriving the maturity levels, the metrics, and the corresponding improvement recommendations". A literature review is already conducted to identify existing DQM maturity models and the CSFs for DQM. The results of this literature review are used to conduct an additional literature review to design and populate the maturity model. The design process of the DQM3 is therefore characterized as 'theory-driven'.

Support of application

As mentioned in section 3.2.4, the existing maturity models for the DQM domain does not provide additional supporting materials that are needed to perform a self-assessment. To tackle this, the DQM3 designed in this study is provided with a 'textual description of form and functioning', and an 'instantiation' in the form of an assessment instrument.

Application method

As defined in the problem statement, the DQM3 should allow an organization to perform a self-assessment in which the DQM maturity level of a business chain is measured. The characteristic for this decision parameter is therefore characterized as 'self-assessment'.

Respondents

The 'respondents' decision parameter is characterized as 'management', 'staff' and 'business partners', because the assessment instrument is designed in the way that it can be completed by a large number of individuals within the unit of analysis. The rationale for this decision can be found in section 4.2.3.

Decision parameter	*Characteristic*			
Maturity definition	Process focused	Object focused	People focused	Combination
Goal function	One-dimensional		Multi-dimensional	
Design process	Theory-driven	Practitioner-based	Combination	
Design product	Textual description of form	Textual description of form and functioning	Instantiation (assessment tool)	
Application method	Self-assessment	Third-party assisted	Certified professionals	
Respondents	Management	Staff	Business partners	Certified professionals

Table 12 - Design model decision parameters and characteristics

4.2.1 Structure of the DQM3

The CDQ MM of Hüner, Ofner, & Otto (2009) divides the concept of maturity models into two generic sub-concepts: a 'domain reference model' (represents the scope or domain that is assessed), and an 'assessment model' **(**defines how certain maturity levels are assigned to elements of the domain reference model). Representing a maturity model in this way "allows constructing exchangeable components and enables easier reuse of already existing components" (Hüner, Ofner, & Otto, 2009). Due to the flexible nature of this structure, these concepts are adopted and reused in the DQM3 (Figure 5).

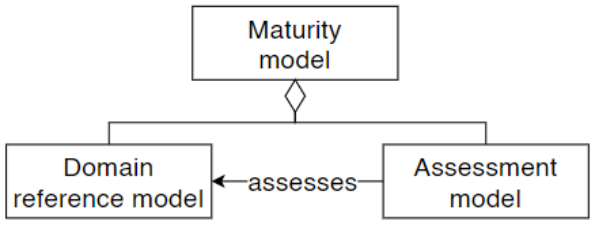


Figure 5 - Components of the DQM3 (Adapted from Hüner, Ofner, & Otto, 2009)

The domain reference model of the DQM3 is represented in a three-tiered layered model as proposed by De Bruin et al. (2005): a domain consisting of multiple domain components, which in turn consists of multiple domain sub-components. Representing a domain reference model in layers "enables an organization to gain a deeper understanding of their relative strengths and weaknesses in the domain and to target improvement strategies thereby enabling more resource allocation" (de Bruin et al., 2005). Figure 6 visualizes the coherence between the three layers and the components.

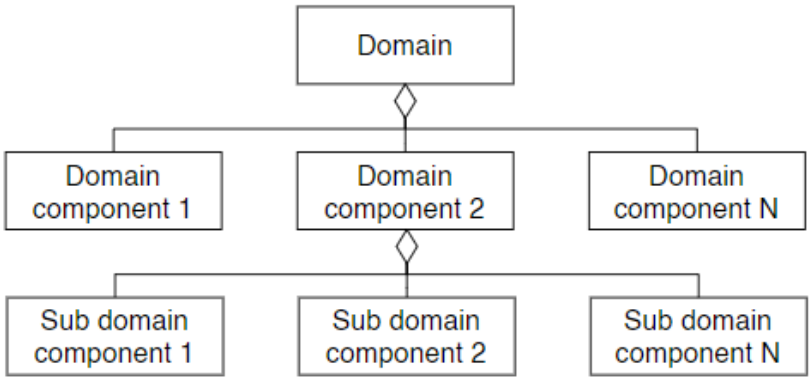


Figure 6 - Components of the domain reference model

4.2.2 Domain reference model of the DQM3

Since the first level of the domain reference model is already defined in the 'define scope' phase, the next step is identify and populate the domain components and sub-components. Rosemann & de Bruin (2004) argued that CSFs provide great insights into such domain components. The CSFs for DQM identified in section 3.1.3 are used as initial input for the domain components. This list is also used as a starting point for the identification of the domain sub-components. Domain sub-components "are specific capabilities that provide further detail enabling targeted maturity level improvements" (de Bruin et al., 2005). For the identification of domain sub-components for the domain reference model, the publication of Santos & Lucas (2019) and Lucas (2019) and the sources they used for their studies are consulted to define the practices.

The publication of Baškarada & Koronios (2014) is one of the most used source for defining the domain sub-components. In this study, inter-dependencies between CSFs for DQM were identified and a CSF framework for DQM was generated that suggests a logical sequence in which those CSFs should be implemented (Baškarada & Koronios, 2014). Although the domain reference model consists of domain components (i.e., merged CSFs from multiple papers), the logical order in this CSF framework and the total quality management implementation three-level framework developed by Hietschold, Reinhardt, & Gurtner (2014) provide relevant information to structure and group the domain components.

Domain reference model – part I

Derived from Baškarada & Koronios (2014), it is suggested that organizations should start their DQM program by addressing the '[MCLS] Management commitment, leadership, and support' and '[DG] Data governance' domain components first since most of other domain components are either directly or indirectly managed by it. This view is supported by the three-level framework developed by Hietschold, Reinhardt, & Gurtner (2014). However, this three-level framework argues that the domain components '[CC] Culture and communication' and '[ET] Education and training' also have an influence on all the other domain components.

The sub-components for the first four domain components that together form the first part of the domain reference model are defined in Table 13. A code is added to each domain sub-component to make future references more convenient.

Domain comp.	Domain sub-component	Practice	References / sources
[MCLS] Management commitment, leadership, and support[3]	**[MCLS-1] Recognition of importance and responsibility**	A. Management should recognize the importance of DQ.	- Wixom & Watson, 2001 - Eckerson, 2002 - Zhang et al., 2003 - Yeoh & Koronios, 2010 - Xu, 2015
		B. Management should take responsibility for the quality of the data.	
	[MCLS-2] Allocation of resources	Management should allocate sufficient resources (budget, technical tools (software), expertise, skilled personnel, and time to support DQ.	- Wixom & Watson, 2001 - Zhang et al., 2003 - Ariyachandra & Frolick, 2008 - Yeoh & Koronios, 2010 - Hietschold, Reinhardt, & Gurtner, 2014 - Xu, 2015 - Santos & Lucas, 2019 - Lucas, 2019

[3] Management is not confined to top or senior management, but refers to all levels of management in the organization (Tee et al., 2007).

	[MCLS-3] Attribution of rewards	Management should attribute rewards to employees who are proactively committed to managing DQ. • Examples of rewards: recognition for DQ improvement suggestions; increased budget; positive feedback.	- Porter & Parker, 1993 - Tee et al., 2007 - Santos & Lucas, 2019
	[MCLS-4] Strategic data quality plan	There should be a clear plan that provides how the DQ will be improved. Additional: • The strategy has to be linked or integrated with the enterprise business strategy. • Employees should be educated about the strategy and their roles in achieving the goals of the strategy.	- Eckerson, 2002 - Akhavan, Jafari, & Fathian, 2006 - Ariyachandra & Frolick, 2008 - Hietschold, Reinhardt, & Gurtner, 2014 - Santos & Lucas, 2019 - Lucas, 2019
[DG] Data governance	**[DG-1] Organizational structure and roles/responsibilities**	The organization should have a suitable organizational structure with clear roles and responsibilities for the DQ to produce high-quality data. Example DQ roles and responsibilities (Eckerson, 2002): • *Data steward/data champion*: a business person who is accountable for the quality of data in a given subject area; • *Subject matter expert*: a business analyst whose knowledge of the business and systems is critical to understand data, define rules, identify errors, and set thresholds for acceptable levels of DQ; • *DQ leader*: oversees a DQ program that involves building awareness; developing assessments, establishing SLAs, cleaning and monitoring data, and training technical staff; • *DQ analyst*: responsible for profiling, monitoring, and measuring DQ on a daily basis, and recommending actions for correcting and preventing errors and defects; • *Tools specialists*: individuals who understand either ETL or DQ tools or both and can translate business requirements into rules that these systems implement. • *Process improvement facilitator*: coordinates efforts to analyze/reengineer business processes to streamline data collection, exchange, and management, and improve DQ. • *DQ trainer*: develops and delivers DQ education, training, and awareness programs.	- Porter & Parker, 1993 - Wixom & Watson, 2001 - Eckerson, 2002 - Xu, Koronios, & Brown, 2003 - Ariyachandra & Frolick, 2008 - Yeoh & Koronios, 2010 - Xu, 2015 - Santos & Lucas, 2019 - Lucas, 2019
	[DG-2] Teamwork	There should be effective teamwork between business and IT people (within and between different teams/departments), as a key to improving DQ.	
[CC] Culture and communication	**[CC-1] Organizational culture focused on data quality**	The organization should have a culture that focuses on the continuous and proactive management of DQ.	- Xu, 2015 - Zhang et al., 2003 - Baškarada & Koronios, 2014 - Hietschold, Reinhardt, & Gurtner, 2014 - Santos & Lucas, 2019 - Lucas, 2019
	[CC-2] Communication	There should be effective communication between business and IT people (within and between different teams/ departments), as a key to improving DQ.	- Xu, Koronios, & Brown, 2003 - Ariyachandra & Frolick, 2008 - Hietschold, Reinhardt, & Gurtner, 2014 - Xu, 2015 - Santos & Lucas, 2019
	[CC-3] Knowledge sharing	Knowledge should be shared at regular/event-triggered knowledge sharing occasions, and via networks of experts. • *Regular* – repeated at specific intervals • *Event-triggered* – at specific events like e.g., a project's end, coming up of a new technology	- Akhavan, Jafari, & Fathian, 2006 - Santos & Lucas, 2019

		• *Networks of experts* – e.g., knowledge teams or centers, scientific committees, communities of practice.	
[ET] Education and training	**[ET-1] Training workshops/ programs**	The organization should ensure that employees are deeply familiar with DQM by offering sufficient initial and ongoing training workshops/programs. • *Initial training*: new personnel, new/upgrade systems • *Ongoing training*: regular training for employees/managers.	- Eckerson, 2002 - Zhang et al., 2003 - Xu, Koronios, & Brown, 2003 - Akhavan, Jafari, & Fathian, 2006 - Baškarada & Koronios, 2014 - Xu, 2015 - Santos & Lucas, 2019 - Lucas, 2019
	[ET-2] Professional development	The organization should support the on-the-job professional development of employees (e.g., by providing mentoring programs).	- Baškarada & Koronios, 2014 - Lucas, 2019

Table 13 - Domain reference model of the DQM3 (part I)

According to Lucas (2019) and highlighted by the '[L-11] Personnel competency' CSF, the competency of personnel who are responsible for DQ is of high importance. Therefore, the organization should "employ well-trained, experienced, and qualified individual personnel at all levels, from top and middle management to employees" (Xu, 2015). However, as shown in Table 13 and defined by the '[MCLS-2] Allocation of resources' domain sub-component, management is responsible for allocating sufficient resources (including expertise and personnel that is highly skilled) to support DQM activities.

De Bruin et al. (2005) argued that "the number of domain components and sub-components should be kept low" when designing a maturity model so that the perceived complexity in the model is minimized and the independence of the components is ensured. Therefore, the decision has been made to further limit the initial list of domain components and represent personnel competency as part of the '[MCLS-2] Allocation of resources' domain sub-component.

Domain reference model – part II

The next group of closely related CSFs that represent so-called 'quality assurance processes' (according to Baškarada & Koronios (2014)) correspond to the following four domain components from the domain reference model: (1) '[DQREM] Data quality requirements management', (2) '[DQRIM] Data quality risk management', (3) '[DQAM] Data quality assessment/monitoring', and (4) '[CDQI] Continuous data quality (management) improvement'. Supported by the second level of the three-level framework developed by Hietschold, Reinhardt, & Gurtner (2014), this group is supplemented with (5) '[DCS] Data customer satisfaction' and (6) [DSQM] Data supplier quality management' domain components since these are both focused on quality assurance processes instead of operational processes.

Despite that both domain sub-components of '[DCS] Data customer satisfaction' are reflected in other domain components, it has been decided to not further limit the amount of domain components by excluding '[DCS] Data customer satisfaction' but to be consistent with the paper of Lucas (2019) in which this domain component was extracted from the initial '[DQREM] Data quality requirements management' domain component.

The sub-components for the domain components that together form the second part of the domain reference model are defined in Table 14. A code is added to each domain sub-component to make future references more convenient.

Domain comp.	Domain sub-component	Practice	References / sources
[DQREM] Data quality requirements management	[DQREM-1] Stakeholder management	A. Data quality should be central to the prioritization of requests that are realized by the organization (e.g., by prioritizing the key stakeholders).	- Baškarada & Koronios, 2014 - Lucas, 2019
		B. The organization should focus on the needs and quality requirements that data customers have (e.g., by enabling active participation to ensure and improve DQ) (+ [DCS-1]).	- Xu, Koronios, & Brown, 2003 - Hietschold, Reinhardt, & Gurtner, 2014 - Xu, 2015 - Santos & Lucas, 2019 - Lucas, 2019
	[DQREM-2] Data (quality) requirements management	A. Data requirements should be accurately captured in conceptual, logical, and physical data models.	- Baškarada & Koronios, 2014 - Lucas, 2019
		B. DQ requirements should be proactively identified, verified, validated, and updated in consultation with stakeholders. • DQ requirements: key DQ dimensions, relevant business rules, and minimum desirable levels of DQ.	
[DQRIM] Data quality risk management	[DQRIM-1] Awareness of consequences	The organization should be aware of the consequences that poor DQ can have (e.g., by proactively identifying key risk areas).	- Xu, Koronios, & Brown, 2003 - Baškarada & Koronios, 2014 - Xu, 2015 - Lucas, 2019
	[DQRIM-2] Commitment to impact reducing	The organization should be committed to reducing the impact that poor DQ can have (e.g., by proactively monitoring and, if possible, mitigating/controlling the key risks).	
[DQAM] Data quality assessment/monitoring	[DQAM-1] Data profiling	Data profiling should be used adequately to proactively identify DQ problems (e.g., missing data, incorrect values, duplicate records, and violations of business rules).	- Eckerson, 2002 - Xu, Koronios, & Brown, 2003 - Ariyachandra & Frolick, 2008 - Yeoh & Koronios, 2010 - Baškarada & Koronios, 2014 - Xu, 2015 - Santos & Lucas, 2019 - Lucas, 2019
	[DQAM-2] Data quality assessment/ monitoring	A. The current state of DQ should be monitored effectively with DQ tools (e.g., SAP Information Steward) (e.g., by creating true DQ assessments on regular intervals). Strict DQ assessment should be performed at source systems (at reception points) and data hubs. The results of these assessments can be used to: 1. Monitor the compliance with policies, standards, service level agreements (SLAs), and legal or regulatory constraints; 2. Measure the progress in achieving DQ goals; 3. Recommend actions for fixing (cleansing) the data;	- Eckerson, 2002 - Xu, Koronios, & Brown, 2003 - Akhavan, Jafari, & Fathian, 2006 - Hietschold, Reinhardt, & Gurtner, 2014 - Baškarada & Koronios, 2014 - Xu, 2015 - Santos & Lucas, 2019 - Lucas, 2019
		B. The organization should assess the subjective perceptions of data quality of the data customers (+ [DCS-2]).	
[DCS] Data customer satisfaction	[DCS-1] Focus on data customers	This practice is reflected in part B of: [DQREM-1] Stakeholder management	
	[DCS-2] Data customer's perception	This practice is reflected in part C of: [DQAM-2] Data quality assessments / monitoring	

[CDQI] Continuous data quality (management) improvement	**[CDQI-1] Continuous improvement of data quality controls** (preventive, detective, and corrective)	The organization should be focused on the continuous improvement of (system and human) preventive, detective, and corrective DQ controls. • *Preventive controls*: to prevent poor quality data from entering the systems (e.g., data validation checks (i.e., input controls), provision of staff training). • *Detective controls*: to detect poor quality data in the systems (e.g., data profiling tools/scripts). • *Corrective controls*: to correct poor quality data in the systems (e.g., data cleansing). E.g.: by monitoring effectiveness of processes, by learning from internal and external sources, by reengineering business processes.	- Eckerson, 2002 - Xu, Koronios, & Brown, 2003 - Zhang et al., 2003 - Akhavan, Jafari, & Fathian, 2006 - Huang, Wu, & Chen, 2013 - Baškarada & Koronios, 2014 - Xu, 2015 - Santos & Lucas, 2019 - Lucas, 2019
	[CDQI-2] Cost-benefit analyzes	Systematic cost-benefit analyzes should be performed before any process improvement are made, to maximize benefits at minimum costs.	- Xu, Koronios, & Brown, 2003 - Baškarada & Koronios, 2014 - Xu, 2015 - Santos & Lucas, 2019 - Lucas, 2019
	[CDQI-3] Change management	The organization should be capable to effectively manage both internal and external changes. • *Internal changes*: e.g., organization restructuring, introducing new technology, personnel changes, culture changes, business process changes. • *External changes*: e.g., government regulations, technology, economy, and market changes.	- Xu, 2015 - Xu, Koronios, & Brown, 2003 - Santos & Lucas, 2019 - Lucas, 2019
[DSQM] Data supplier quality management	**[DSQM-1] Agreement about acceptable data quality level**	This practice is reflected in: [DPLM-1] Data product supply chain management	- Xu, Koronios, & Brown, 2003 - Xu, 2015 - Lucas, 2019
	[DQSM-2] Data quality reports and technical assistance	External data suppliers should be provided with regular DQ reports and technical assistance.	

Table 14 - Domain reference model of the DQM3 (part II)

The 'evaluate cost/benefit tradeoffs' domain component was merged from two CSFs that have same descriptors. According to Baškarada & Koronios (2014), it is critical to perform cost-benefit analysis before any process improvements are made. Therefore, in line with Baškarada & Koronios (2014), 'Cost-benefit analyzes' is integrated in '[CDQI] Continuous data quality (management) improvement' as domain sub-component.

The same applies for the 'change management' domain component. This highlights that an organization should be able to manage both internal and external changes, taking into account the DQ requirements of these changes. To further limit the domain components, the decision has been made to integrate this in '[CDQI] Continuous data quality (management) improvement' as domain sub-component.

Domain reference model – part III

The three domain components that correspond to the CSFs that are part of the other group with closely related CSFs are the following: (1) '[DPLM] Data product lifecycle management', (2) '[DAM] Data architecture management', and (3) '[DSM] Data security management'. This group represents so-called 'operational processes'.

Table 15 shows the identified sub-components for the 'operational processes' domain components that together form part III of the domain reference model. A code is added to each domain sub-component to make future references more convenient.

Domain comp.	Domain sub-component	Practice	References / sources
[DSM] Data security management	[DSM-1] Access controls and security	A. Users of information systems should be appropriately authenticated as well as authorized with the least set of privileges they require.	- Xu, Koronios, & Brown, 2003 - Baškarada & Koronios, 2014 - Xu, 2015 - Santos & Lucas, 2019 - Lucas, 2019
		B. Audit trails (logs of users' activities) of the information systems used should be analyzed and periodically reviewed (e.g., for exceptions).	
	[DSM-2] Human and process controls	Segregation of duties should be ensured (i.e., having more than one person required to complete a critical task to prevent fraud and/or error).	
[DPLM] Data product lifecycle management	[DPLM-1] Data product supply chain management	There should be clear agreements about the acceptable DQ levels (e.g., by establishing Service Level Agreements (SLAs).	- Baškarada & Koronios, 2014 - Eckerson, 2002 - Lucas, 2019
	[DPLM-2] Data migration	A. Data migration (i.e., Extract Transform Load (ETL) should be minimized as much as possible (e.g., only undertaken in the case of system/database upgrades).	- Baškarada & Koronios, 2014 - Santos & Lucas, 2019
		B. Standard methods for system/software development should be followed and extensive testing should be included when migrating data (i.e., ETL).	
	[DPLM-3] Metadata management	Metadata should be appropriately managed (e.g., by organizing it in a metadata repository, separately from the transactional and master data).	- Baškarada & Koronios, 2014 - Eckerson, 2002
[DAM] Data architecture management	[DAM-1] Data integration	Integration of data (i.e., combining data from different sources into a single, unified view) should be maximized as much as possible (e.g., by implementing a data warehouse).	- Eckerson, 2002 - Tee et al., 2007 - Ariyachandra & Frolick, 2008 - Santos & Lucas, 2019
	[DAM-2] Interface management	Interfaces (i.e., links between different systems) should be minimized as much as possible.	- Eckerson, 2002 - Baškarada & Koronios, 2014 - Santos & Lucas, 2019

Table 15 - Domain reference model of the DQM3 (part III)

Limiting the domain components

Besides the opportunities to further limit the list of domain components by integrating domain components as sub-components in other domain components, some problems arose while developing the domain reference model. These problems causes the need to exclude the following domain components:

- **Documentation**. Santos & Lucas (2019) defines 'documentation' as CSF for DQM and provides four references as sources to identify the sub-components for this domain component. However, useful information could only be extracted from the publication of Xu, Koronios, and Brown (2003): "adequate and sufficient documentation for people to follow".

  This was defined under 'nature of the IS' that is previously classified as a contingency factor by the experts and the decision was made to not include those factor to limit the amount of domain components. Therefore, this domain component is not included in the domain reference model.

- **Data quality policies and standards**. In both the study of Lucas (2019) and Santos & Lucas (2019), 'data quality policies and standards' is identified as a CSF for DQM. However, the only relevant information that could be extracted from the sources that are used by Lucas (2019) and Santos & Lucas (2019) to identify the sub-components for this CSF is that according to Xu (2015) and Xu, Koronios, & Brown (2003), an organization should have in place "appropriate (simple, relevant, and consistent) DQ policies and standards" by establishing and implementing/enforcing those policies and standards. Besides this, it is not defined what exactly is meant by 'policies and standards'. Nevertheless, monitoring compliance with DQ policies and standards is part of '[DQAM-2] Data quality assessments/ monitoring', so it is decided to exclude this domain component as an independent component in the domain reference model. The role of this CSF will need to be further explored in future research.

Domain reference model – overview

As identified in the previous subsections, the domain reference model is populated with 13 domain components and 31 domain sub-components. To summarize, the final list of these domain components and domain sub-components are shown in Table 16.

Domain component	Domain sub-component
[MCLS] Management commitment, leadership, and support	[MCLS-1] Recognition of importance and responsibility
	[MCLS-2] Allocation of resources
	[MCLS-3] Attribution of rewards
	[MCLS-4] Strategic data quality plan
[DG] Data governance	[DG-1] Organizational structure and roles/responsibilities
	[DG-2] Teamwork
[CC] Culture and communication	[CC-1] Organizational culture focused on data quality
	[CC-2] Communication
	[CC-3] Knowledge sharing
[ET] Education and training	[ET-1] Training workshops/programs
	[ET-2] Professional development
[DQREM] Data quality requirements management	[DQREM-1] Stakeholder management
	[DQREM-2] Data (quality) requirements management
[DQRIM] Data quality risk management	[DQRIM-1] Awareness of consequences
	[DQRIM-2] Commitment to impact reducing
[DQAM] Data quality assessment/monitoring	[DQAM-1] Data profiling
	[DQAM-2] Data quality assessment/monitoring
[DCS] Data customer satisfaction	[DCS-1] Focus on data customers
	[DCS-2] Data customer's perception
[CDQI] Continuous data quality (management) improvement	[CDQI-1] Continuous improvement of data quality controls (preventive, detective, and corrective)
	[CDQI-2] Cost-benefit analyzes
	[CDQI-3] Change management

[DSQM] **Data supplier quality management**	[DSQM-1] Agreement about acceptable data quality level
	[DSQM-2] Data quality reports and technical assistance
[DSM] **Data security management**	[DSM-1] Access controls and security
	[DSM-2] Human and process controls
[DPLM] **Data product lifecycle management**	[DPLM-1] Data product supply chain management
	[DPLM-2] Data migration
	[DPLM-3] Metadata management
[DAM] **Data architecture management**	[DAM-1] Data integration
	[DAM-2] Interface management

Table 16 - Domain reference model of the DQM3

4.2.3 Assessment model of the DQM3

With the domain reference model constructed and populated with domain components and sub-components, the next step is to determine how maturity measurement can take place, i.e., to develop the assessment model of the DQM3 (de Bruin et al., 2005). This includes the development of the instantiation that will be used when performing a self-assessment (as defined by the decision parameter 'application method') and the inclusion of appropriate questions and measures within this instrument.

De Bruin et al. (2005) recommend a quantitative method such as a survey that can be made available through electronic means (also referred to as 'online survey' by Sue & Ritter (2007)) for doing a maturity assessment. Two major advantages of such online surveys are (de Bruin et al., 2005):

1. Easily distributable to a wide range of respondents, also across geographic boundaries; and
2. Reduced costs associated with survey distribution.

The instrument required to conduct an online survey is an online questionnaire. According to Sue & Ritter (2007), there are two types of online questionnaires: (1) 'web-page questionnaires' and (2) 'e-mail questionnaires'. The web-page type of questionnaire is selected for the assessment model because these "include a wide variety of question types and can be programmed" in such a way that questions can be skipped when necessary (Sue & Ritter, 2007). Web-page questionnaires also ensure "more accurate data than when respondents are asked to skip questions in an e-mail questionnaire" (Sue & Ritter, 2007).

Assessment instrument: structure of the online questionnaire

Based on the structure of the domain reference model, the domain components and sub-components, and using the benefits of an online questionnaire, the assessment instrument of the DQM3 consists of five main parts and has the following structure:

Part I *Welcome*

It is important to introduce an online questionnaire "with a welcome screen that is motivational, emphasizes the ease of responding, and instruct respondents on the action needed for proceeding to the next page" (Sue & Ritter, 2007). This screen will be used to describe the purpose of the survey, to explain why the respondent was selected for participation, and to discuss the conditions of confidentiality and anonymity. In addition to these elements proposed by Sue & Ritter (2007), information about the business chain can be added to clarify to the respondents where the scope of the business chain has been defined.

It is possible to structure an online questionnaire in two different ways: (1) in discrete blocks with each block being completed by experts in the component" or (2) in such a way that the questionnaire will be "completed by a large number of individuals within the unit of analysis" (de Bruin et al., 2005). Designing the survey in the first way, the number of respondents will be limited and therefore the ability to generalize within the unit of analysis decreases. However, the second way increases the ability to generalize within the unit of analysis but results in a less comprehensive understanding of the domain components because a fewer questions with a more general intent are asked.

Part II Contingency question

Based on the domain reference model, the decision has been made to increase the ability to generalize within a business chain by designing the questionnaire in a way that all individuals who are involved in that business chain can complete the questionnaire. However, to ensure that respondents are not forced to read and/or answer questions that are not meant for them, a contingency question is added to the assessment instrument. A contingency question "allow your respondents to be redirected to a new set of questions on the basis of their responses" (Sue & Ritter, 2007). The contingency question will split the respondents in the following three groups that are identified from the domain reference model: (1) internal data suppliers, (2) external data suppliers, and (3) data customers. These groups are used to define logic at domain component level or sub-component level.

Part III *Department/ team*

Designing the questionnaire in the way that it can be "completed by a large number of individuals within the unit of analysis" results in a less comprehensive understanding of the domain components. To gain a more comprehensive understanding, the organization should identify which departments/teams (depends on how the organization is organized) are involved in the business chain and add them to the questionnaire under the right contingency group. The answer selected on this question will be used as variable within the instrument.

Just as the CSFs provided good guidance for the identification of the domain components and the domain sub-components, these components could subsequently provide good guidance for the inclusion of appropriate questions and measurement criteria. According to De Bruin et al. (2005) and supported by Sue & Ritter (2007), a review of existing literature should result in a comprehensive list of questions. However, no questions are identified in the studied literature when developing and populating the domain reference model. To address this issue, the questions that are part of this assessment model are formulated based on the domain reference model and knowledge gained during this study.

Part IV *Questions*

According to De Bruin et al. (2005) and Sue & Ritter (2007), it is important to balance the number of questions included in the assessment instrument. Sufficient questions are required to ensure a complete measurement. However, too many questions may reduce the reliability of the collected data because this results in a reduction in total responses to the assessment instrument or an increase in incomplete responses (de Bruin et al., 2005). Therefore, the decision has been made to include only one question per domain sub-component and not formulating additional general questions related to the domain components in general to balance the number of questions as recommended by De Bruin et al. (2005). Only if the practice of the sub-component is divided into multiple parts (e.g., '[DSM-1] Access controls and security'), the numbers of questions can match to this.

Since no questions could be identified form the studied literature, the following four levels of measurement are considered when developing the questions: nominal data, ordinal data, 'interval data, and 'ratio data (Sue & Ritter, 2007). According to De Bruin et al. (2005), the most popular way is representing maturity on a five-point Likert scale with '5' representing the highest level. The application of these Likert scales correspond to the 'ordinal data' level of measurement. Using Likert scales can also improve the reliability and consistency of responses and enables results to be easily mapped to maturity levels (de Bruin et al., 2005). Therefore, Likert scales are used as answer options for every question included in this assessment instrument by applying or adapting the Likert-type scale response anchors of Vagias (2006).

Part V *Thank you*

In the final part of the questionnaire, the respondents are thanked for participating in the survey. According to Sue & Ritter (2007), this will reassure for the respondent that his or her opinion will be taken seriously.

These five parts together form the assessment instrument of the DQM3 and is designed in Qualtrics Survey Software. A blueprint of this assessment instrument can be found in Appendix III. The link to the online questionnaire with automated logic and can be found in the next chapter. The online version is both in Dutch and English.

Maturity scores, maturity levels, and example calculation

To finalize the assessment model of the DQM3, five maturity levels with general descriptors are defined (Table 17). The decision was made not to adopt any of the descriptors from the existing DQM maturity models since these do not reflect the complex situation in the domain reference model and the use of Likert scales as measurement criteria for all questions in the online questionnaire.

As shown in Appendix III, maturity scores are assigned to the measurement criteria of all questions. These maturity scores are implemented to the Qualtrics Survey Software back-end and are automatically generated by the respondents when completing the questionnaire. Such a maturity score belongs to the same maturity level, i.e., maturity score '2' corresponds to the 'Level 2 – Low DQM maturity'. An example calculation of maturity scores and levels is presented in Table 18.

Maturity levels, descriptors, and criteria	
	When to assign this maturity level to a domain sub-component / domain component / domain? If average maturity score is…
Level 1 – Very low DQM maturity	Below 2
Level 2 – Low DQM maturity	>= 2 or < 3
Level 3 – Moderate DQM maturity	>= 3 or < 4
Level 4 – High DQM maturity	>= 4 or < 5
Level 5 – Very high DQM maturity	5

Table 17 - Maturity levels, descriptors, and criteria

Example calculation of maturity scores and levels on domain sub-component, domain component, and domain level

There are two respondents: *respondent A* and *respondent B*. The domain reference model consists of two domain components (X and Y) and four domain sub-components: evenly distributed over the domain components (X1, X2, Y1, Y2).

Question	Respondent A answers	Respondent B answers		Maturity score A	Maturity score B
X1	Strongly disagree	Disagree	>	1	2
X2	Very committed	Very committed		4	4
Y1	Somewhat effective	Not effective at all		3	1
Y2	Agree	Strongly agree		4	5

Maturity level of domain sub-component:

- Average maturity score of X1 = 1 + 2 = 3 / 2 = 1.5 > Level 1 – Very low DQM maturity
- Average maturity score of X2 = 4 + 4 = 8 / 2 = 4 > Level 4 – High DQM maturity
- Average maturity score of Y1 = 3 + 1 = 4 / 2 = 2 > Level 2 – Low DQM maturity
- Average maturity score of Y2 = 4 + 5 = 9 / 2 = 4.5 > Level 4 – High DQM maturity

Maturity level of domain component:

- Average maturity score of X = 1.5 + 4 = 5.5 / 2 = 2.75 > Level 2 – Low DQM maturity
- Average maturity score of Y = 2 + 4.5 = 6.5 / 2 = 3.25 > Level 3 – Moderate DQM maturity

Maturity level of domain:

- Average maturity score of DQM = 2.75 + 3.25 = 6 / 2 = 3 > Level 3 – Moderate DQM maturity

Table 18 - Example calculation of maturity scores and levels

To summarize, the relation between the domain reference model and the assessment model is visualized in Figure 7. A question assesses a certain domain sub-component. A response is generated when answering a question. A maturity score is assigned to a response and represents one of the five maturity levels. For the domain and every domain and sub-component, a maturity level is calculated by taking the mean of all respondents.

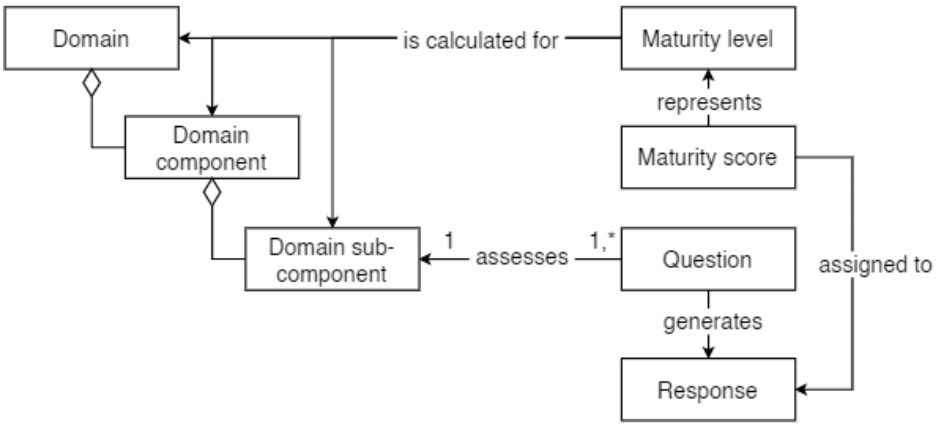


Figure 7 - Relation between the domain reference model and the final assessment model

5. Evaluate Design (Treatment Validation) & Reflect Evolution

***SQ4**: How is the designed maturity model used in practice?*

This chapter presents the results of the final two phases of the development cycle: evaluate design, and reflect evolution (Mettler, 2011). Section 5.1 discusses the results of a pretest and a case study conducted at Achmea. In section 5.2, the design mutability of the DQM3 is contemplated.

5.1 Evaluate design

Mettler (2011) defined the following three decision parameters that help to make critical decisions for the 'evaluate design' phase: subject of evaluation, time-frame, and evaluation method. The decisions made for these three parameters are discussed below and summarized in Table 19.

Subject of evaluation

According to Mettler (2011), it is possible to test two elements: (1) the design process (i.e., the way the model was constructed), or (2) the design product (i.e., the maturity model itself). Mettler's (2011) opinion is that both should be subject to evaluation. However, due to time constraints, limited resources, and the business need of Achmea to perform a maturity assessment, only the 'design product' (DQM3) is subject to evaluation.

Time-frame & evaluation method

As this study focuses on conducting a case study at Achmea, the decision for the 'time-frame' and 'evaluation method' parameters is already defined. The timeframe is characterized as 'ex-post' and the evaluation method is characterized as 'naturalistic' by performing two activities. First, the assessment instrument is reviewed for content and technical difficulties during a pretest. Second, a case study is conducted to evaluate the maturity model on its completeness and validity.

Decision parameter	*Characteristic*		
Subject of evaluation	Design process	Design product	Both
Time-frame	Ex-ante	Ex-post	Both
Evaluation method	Naturalistic	Artificial	

Table 19 - Evaluate design parameters

5.1.1 Pretest of the assessment instrument

According to De Bruin et al. (2005), "it is important that questions and responses are valid, i.e., that they measure what it is they are intended to measure." To test this before actually performing a maturity assessment, Sue & Ritter (2007) propose "to pretest the questions on a small sample of your target population to ensure that respondents understand the scale labels the way you indented them to be understood." Such a pretest "will provide feedback about the ease of navigation, and an understanding of the target population will aid in the inclusion of items that are interesting and relevant to the respondents" (Sue & Ritter, 2007).

To ensure that the questions measure what it is they are intended to measure, the assessment instrument has been reviewed for content and technical difficulties in such a pretest, as proposed by Sue & Ritter (2007). A sample of two participants from the target population of the case study (a team manager and a business analyst, both from the 'Datalogistiek' team) were selected to complete the online questionnaire. These employees were asked during an online session to provide feedback on (1) whether the

components contribute to DQM maturity, and (2) both the content difficulties (clarity of the questions, and whether the initial questions were formulated at the right level) and technical difficulties (i.e., the proper functioning of the technical elements) they experienced while completing the online questionnaire. The results of this pretest are discussed below.

Results of the pretest and improvement made to the assessment instrument

Overall, the assessment instrument performed well during the pretest. The two participants experienced no technical difficulties while completing the online questionnaire. Both participants also reacted positively on the components (and questions) included in the assessment instrument and confirmed that those indeed influence DQM and recognized the importance of these components. However, there were some content difficulties regarding the clarity of some questions. The participants also indicated that some questions could be formulated at a different level. In consultation with the participants of the pretest, the following improvements are made to the assessment instrument (Table 20):

Sub-component	*Content difficulties and improvements*
[MCLS-2] Allocation of resources	The statements on 'sufficient expertise' and 'sufficient skilled personnel' were a bit unclear. To tackle this unclearness, 'sufficient expertise' is changed to 'sufficient expertise (knowledge)'. Also, 'sufficient skilled personnel' is changed to 'sufficient personnel (skills)'.
[MCLS-4] Strategic data quality plan	To gain a more comprehensive understanding, the participants suggested adapting the statement by using the team/department level instead of the business chain level. This suggestion is accepted.
[DG-1] Organizational structure and roles/responsibilities	To gain a more comprehensive understanding, the participants suggested dividing the question into two statements. This suggestion is accepted.
[DG-2] Teamwork	The information between brackets ('within and between different teams/departments') is deleted. This is because this information is already integrated into the statements.
[CC-2] Communication	The information between brackets ('within and between different teams/departments') is deleted. This is because this information is already integrated into the statements.
[CDQ-1] Continuous improvement of data quality controls	To gain a more comprehensive understanding, the participants suggested adapting the questions and corresponding statements by using the team/department level instead of the organizational level. This suggestion is accepted.
[CDQI-4] Change management	The statements on 'managing internal changes' and 'managing external changes' were a bit unclear. Examples of internal and external changes are included within the statements to tackle this unclearness.
[DPLM-3] Metadata management	To gain a more comprehensive understanding, the participants suggested adapting the statement by using the team/department level instead of the organization level. This suggestion is accepted.

Table 20 - Content difficulties and improvements made to the assessment instrument

In addition to the feedback given on the formulation of questions, the two participants also questioned whether the measurement criteria provide enough possibilities for all respondents to answer the questions. They indicated that this would mainly be an issue with the questions that are formulated on the 'business chain' level (e.g., the questions on '[DAM-1] Data integration' and '[DAM-2] Interface management'). The current questionnaire does not provide an opportunity for respondents to indicate that he or she has a lack of knowledge about a specific sub-component.

To tackle this issue, a sixth answer option ('not applicable (N/A) /no idea') will be added to the Likert scales of all questions during the case study (see section 5.1.2.2). This answer option could provide valuable information for the organization because a high response to this answer suggests that respondents are not aware of what is going one along the entire business chain, but it could also indicate that the concerned sub-component does not fit in the domain reference model of the DQM3.

Revised assessment instrument

Based on feedback provided during the pretest, improvement are made to the assessment instrument is revised (Table 20). The blueprint of the revised assessment instrument can be found in Appendix IV. Improvements are indicated in red text.

5.1.2 Case study at Achmea

For the case study, a protocol is constructed using the template from Brereton et al. (2008). This protocol is based on research done by Yin (2003) who states that using a protocol improves the reliability. From the template of Brereton et al. (2008), only the following relevant items are discussed: design, data collection, data analysis, and validity. The other items (case selection, case study roles, schedule, study limitations, and reporting) have been previously defined or will be discussed in later sections of this thesis report.

5.1.2.1 Design

For the design of the case study, there were four different types of case study designs to consider (Yin, 2003): holistic single-case design, embedded single-case design, holistic multiple-case design, and embedded multiple-case design. This case study regard a holistic single-case design as it contains one case to be examined: Achmea, and one unit of analysis: 'the data logistic chain (Datalogistiek)' that consists of all teams responsible for the data flow from source systems to internal/external customers. Despite the goal of making the DQM3 fit for other organizations as well, which would imply multiple case studies, this case study is only conducted at Achmea due to time constraints and limited resources. The 'object of study' of this case study is the DQM3 developed in the previous chapter.

5.1.2.2 Data collection

No existing data collection procedures are identified in the publications of the DQM maturity models or other literature studied for the creation of the DQM3. Therefore, based on the domain reference model and the assessment model, the following data collection procedure has been developed and used in this case study:

1. Identify teams/departments involved in the business chain and place them under the right contingency group

As defined in section 4.2.3, the online questionnaire is designed in such a way that it can be completed by a large number of individuals within the business chain. The first step was to identify the teams/departments involved in the business chain and add them under one of the three contingency groups: internal data supplier, external data supplier, and data customer. The following teams involved in 'the data logistic chain (Datalogistiek)' were identified for this case study (Table 21):

<table>
<tr><th>Internal data supplier</th><th>External data supplier</th><th>Data customer</th></tr>
<tr><td rowspan="2">CAD, LiFE, MIAZ, UNiTED, Unicorn, ZZTop, Brainiacs, Fireworks, Run Inkomen, Alice in Legacyland, KeRBI</td><td rowspan="3">SIEBEL</td><td>Marketing Intelligence, Customer Intelligence, CCR</td></tr>
<tr><td rowspan="2">DWH Distributie, CIV, MI4U, FA4U, Marketing & Distributie, Mijn Verzekeringen, Financiële Coach</td></tr>
<tr><td>dP&L BI & Data, Bismuth, Tink, Obelix, BI Joe, Fier, Datalogistiek</td></tr>
</table>

Table 21 - Teams/departments identified for the case study

2. Distribute the assessment instrument to all individuals that are part of the identified teams/departments via email or with an anonymous survey link to collect the individual maturity scores

The assessment instrument was distributed via email to the 'internal data supplier' teams and with an anonymous survey link to the 'external data supplier' and 'data customer' teams. Three modifications have been made to the assessment instrument to collect additional data for evaluation purposes.

First, as proposed during the pretest, a sixth option ('not applicable (N/A) /no idea') was added to the Likert scales of all questions. The respondent was able to select this answer if he/she did not know the answer or if the sub-component was not applicable to the 'the data logistic chain (Datalogistiek)'. A significant response on this option could indicate that a concerned sub-component does not fit in the domain reference model of the DQM3.

Additionally, the following three open questions were added at the end of the online questionnaire regarding the DQM3 and DQM in general to collect additional information used for evaluating the DQM3, drawing conclusions, and making recommendations for improving DQM maturity within 'the data logistic chain (Datalogistiek)':

1. Which components do you think are important for managing data quality but were not covered in this survey?
2. Concerning your role in the chain, what do you think is the biggest challenge in improving data quality management within 'the data logistic chain'?
3. On a scale from 1 (low) to 5 (high), how 'mature' do you think is 'the data logistic chain' in data quality management?

Besides the questions to generate maturity scores for every sub-component, the respondents were also asked to give comments at the end of each domain component block to substantiate their chosen answers. This space for comments was included to provide Achmea with descriptive data outside the boundaries of the theoretical design. This data is also used to determine if sub-components does not fit in the domain reference model or if there were any problems with the question interpretation.

3. Assign an overall maturity level at domain, domain component, and sub-component level and plan improvements

The final step was to calculate the maturity levels based on the individual maturity scores of all respondents. Table 18 shows an example calculation. The calculated levels and notes provided by the respondents are extremely valuable in understanding the current state of DQM maturity in the business chain.

5.1.2.3 Data analysis

The individual scores of the 128 respondents (108 internal data suppliers, 2 external data supplier, and 18 data customers) were automatically stored in Qualtrics. To calculate the maturity levels at all three levels of the domain reference model, the individual scores of the respondents were exported and stored in a spreadsheet. After calculating the maturity levels, the maturity levels were visualized in a table to give the stakeholders a clear overview of the DQM maturity of 'the data logistic chain (Datalogistiek)'.

The results of the maturity assessment at Achmea are not published in this thesis report because of sensitive information that could be derived from the results. These are reported in an external document that is confidential and can only be accessed by Achmea. However, the steps performed to perform the maturity assessment are explained extensively in the previous subsection. The DQM3 is also evaluated on its completeness and validity, the findings of which are discussed in the remainder of this chapter.

Comments on the assessment instrument

As with the pretest, the respondents experienced no technical difficulties while completing the online questionnaire in the case study. However, a few respondents had some comments regarding the assessment instrument in general. One of the respondents found the questionnaire 'to be quite extensive, perhaps a bit too much. Filling out takes a lot of time". This respondent completed the online questionnaire via the 'internal data supplier' flow. Although this flow consists of the most questions (28 in total), the completion rate for this flow was high (144 surveys started, 108 responses, completion rate of 75%). A possible solution to the extensiveness of the questionnaire is to exclude sub-components that are only supported by one reference. However, the identified components that are in the domain reference model seem all to be of high importance since they are all supported by two or more references. The omission of specific sub-components, and therefore also certain questions, will therefore not be applied.

Also, some comments were made on the concept of DQ itself by two respondents that also completed the 'internal data supplier' flow. One respondent indicated that it is important to "first clearly state what is meant by 'data quality'. I know that most of my team members have no idea what this survey is about". The other respondents stated the following: "It is good to explain what exactly is meant by data quality. Everyone may have a different perception of what quality is". However, there is no reason to make such clarification within the assessment instrument, because it is the responsibility of the organization to ensure that there is a common understanding on the definition of DQ (e.g., by providing a training workshops on DQ and DQM).

Besides some comments on the assessment instrument, the case study also showed that the space for comments at the end of each domain component block (added to provide Achmea with descriptive data outside the boundaries of the theoretical design) was beneficial in identifying potential improvement areas within 'the data logistic chain (Datalogistiek)'. This data is used by Achmea to plan specific improvements.

Issues with question interpretation

Despite the pretest to identify all content difficulties and the improvements made to the assessment instrument, two additional issues with question interpretation emerged in the case study. The first issue that appeared was that a respondent, who completed the 'internal data supplier' flow, indicated that he or she was having difficulty determining whether the question on '[CC-3] Knowledge sharing' was a generic question or whether it was explicit about DQ. To avoid interpretation problems with this question in future assessments, the question on '[CC-3] Knowledge sharing' is revised. No evidence is found in the literature that knowledge sharing should specifically focus on DQ, so '(not limited to data quality)' is added at the end of the question.

A second issue emerged regarding the question on '[DSM-2] Human and process controls'. Feedback provided by respondents indicated that multiple respondents focused only on 'segregation of duties' and interpreted this in a way that employees should perform certain activities or actions alone instead of having more than one person required to complete a critical task. Therefore, the question is on '[DSM-2] Human and process controls' changed to avoid interpretation problems with this question in future maturity assessment: more emphasis is placed on the definition of 'segregation of duties', where the term itself is removed from the question.

Missing components

The respondents were asked which components they believe are important for managing DQ, but were not covered in the current model. The following aspects were mentioned by the respondents (Dutch is translated to English), ordered by the number of times mentioned (Table 22):

Times mentioned	*Missing components*	*Additional comments on missing aspect*
3	**Ownership of data**	"Clearly appointed data owners."
3	**Human**	"The 'human' factor may have been a little underexposed; although it is also mentioned."
2	**Tooling**	"Good tooling for capture and easy use." "Good tooling (especially front-end)."
2	**Master data management**	"It starts at the source."
1	**Testing**	"Testing is very important and hardly occurs in this survey."
1	**Time**	"People must have time to manage data quality correctly (and to ensure any knowledge transfer)".
1	**Budget**	x
1	**Input from customers**	"The customer determines how the data should be filled."

Table 22 - Missing components in the domain reference model

The missing component that is mentioned three times by the respondents is 'ownership of data'. A respondent noted that clearly appointed data owners are essential for DQM. This missing component indirectly relates to the already existing sub-component '[DG-1] Organizational structure and roles/responsibilities', which highlights that the roles and responsibilities for the DQ should be clear. If it is not clear within an organization who owns specific data, this suggests that roles and responsibilities are not clear.

The second and third missing components are 'human' and 'tooling'. A respondent highlighted that the 'human' factor might have been a little underexposed, although it is mentioned. As the respondent indicated, this component is already pointed out, namely at '[MCLS-1] Allocation of resources': sufficient expertise (knowledge) and sufficient skilled personnel (skills). This also applies to the missing 'tooling' component: sufficient technical tools (software). Future research on these components can investigate how the 'human' and 'tooling' can be further integrated into the model.

The fourth missing component addressed by two respondents is 'master data management' (MDM). According to Spruit & Pietzka (2015), MDM is "the management of the consistent and uniform subset of business entities that describe the core activities of an enterprise". This maturity model is currently focused on managing the quality of data in general, not only specified to master data. Whether it is necessary to distinguish between master data and other types should be further investigated in future research.

The fifth missing component is 'testing'. A respondent commented that "testing is very important and hardly occurs in this survey". As this comment shows, this component is already included in the current model. The importance of testing is addressed by the '[DPLM-2] Data migration' sub-component: extensive testing should be included when migrating data (i.e., ETL). Whether the role of testing should be further expanded in the DQM3 should be explored in future research.

The sixth and seventh missing components are 'time' and 'budget'. These components are both already included in the current model at '[MCLS-1] Allocation of resources': sufficient time and sufficient budget. The same applies to the missing component 'input from customers' that was brought up by a respondent who completed the questionnaire as an 'internal data supplier'. This component is already part of the model ('[DCS-1] Focus on data customers' that defines than an organization should be focused on the needs and quality requirements that customers have by, e.g., enabling active participation to ensure and improve DQ) but the question related to the component is already part of the 'data customers' flow.

Due to the mostly positive reactions that the respondents do not miss any components in the current domain reference model and the lack of support for the missing components (because most of them are already related (directly or indirectly) to components in the current model or were not identified by the extensive literature review), no new components are added to the domain reference model based on these responses.

Response on the sixth answer option: 'not applicable (N/A) / no idea'

The respondents were able to answer the majority of questions included in the assessment instrument, with some exceptions. Out of 4812 times an answer was selected, 1207 times the sixth answer option ('N/A / no idea') was chosen (25,1%). Regarding the validity of the components, this sixth option was chosen more often (than the overall average of 25,1%) at the questions on the following 12 components: [MCLS-3] Attribution of rewards, [ET-1] Training workshops/programs, [DQAM-1] Data profiling, [DQAM-2a] Data quality assessment/monitoring, [CDQI-2] Cost/benefit analyzes, [CDQI-3] Change management, [DSM-1b] Access controls and security, [DPLM-1] Data product supply chain management, [DPLM-2] Data migration, [DPLM-3] Metadata management, [DAM-1] Data integration, and [DAM-2] Minimization of interfaces. The exact percentages for those components cannot be specified due to the sensitive information that could be derived from them.

Despite the high response to this sixth answer option, both participants of the pretest reacted positively on the components (and questions) included in the assessment instrument and confirmed that those indeed influence DQM and recognized the importance of these components. The comments left by the respondents at the end of each domain component block also mainly showed that the lack of knowledge about the business chain was the reason for choosing the sixth answer option. It explains the high response on this option on the following four components: [DPLM-1] Data product supply chain management, [DPLM-2] Data migration, [DAM-1] Data integration, and [DAM-2] Minimization of interfaces. This is also in line with the concerns of the two participants in the pretest. The other components could be related to the way in which the organization is organized and functions.

Summary of revisions made to the maturity model

Based on the analysis of the data, no improvements are made to the domain reference model of the DQM3. The assessment instrument, on the other hand, is revised for a second time as discussed previously. The blueprint of the final assessment instrument can be found in Appendix V. The actual assessment instrument implemented in Qualtrics can be downloaded via the following link. To summarize, the following improvements are made to the assessment instrument of the DQM3:

Sub-component	*Improvements made*
[CC-3] Knowledge sharing	How often is knowledge shared within [*the organization*] at knowledge sharing occasions and via networks of experts? **… is changed to …** How often is knowledge shared within [*the organization*] at knowledge sharing occasions and via networks of experts? (not limited to data quality)

[DSM-2] Human and process controls	Segregation of duties is ensured within [*your team(s)/department(s)*], i.e., having more than one person required to complete a critical task to prevent fraud and/or error. **… is changed to …** Having more than one person required to complete a critical task to prevent fraud and/or error is ensured within [*your team(s)/department(s)*].

Table 23 - Revisions made to the DQM3

5.1.2.4 Validity

In terms of validity, the following four tests (also referred to as concerns) are often used to determine the quality of empirical research: construct validity, internal validity, external validity, and reliability (Yin, 2003; Brereton et al., 2008). How these tests and its case study tactics are integrated into this case study can be found in Table 24 below.

Test / Concern	*Case study tactic*	*Phase in which tactic occurs*	*Done in this study*
Construct validity: establishing correct operational measures for the concepts being studied.	Use multiple sources of evidence	Data collection	*Not done in this case study (explanation below table)*
	Establish chain of evidence	Data collection	See Figure 8
	Have key informants review draft case study report	Composition	Two participants reviewed the assessment results
Internal validity	*x*		
External validity	*Not done in this case study → future research*		
Reliability: demonstrating that the operations of a study can be repeated with the same results.	Use case study protocol	Data collection	See section 5.3.2 for data collection protocol used
	Develop case study database	Data collection	Maturity scores and notes produced by respondents

Table 24 - Case study validity and reliability concerns

Construct validity

According to Yin (2003), a researcher must be sure to cover the following steps to meet the test of construct validity: "select the specific types of changes that are to be studied (and relate them to the original objectives of the study), and demonstrate that the selected measures of these changes do indeed reflect the specific types of change that have been selected." Using multiple sources of evidence, establishing a chain of evidence, and having key informants review draft case study reports are the case study tactics that increases construct validity:

1. *Use multiple sources of evidence*. The first case study tactic is using multiple sources of evidence. This tactic is relevant during data collection. This case study tactic is not applied in this case study. However, because the questions were already tested in the pretest for clarity and whether the questions were asked at the right level discussed in section 5.1.1, multiple sources of evidence are indirectly used in the treatment validation task.

2. *Establish chain of evidence.* Establishing a chain of evidence increases construct validity, reliability, and allows external observers of a study "to follow the derivation of any evidence, ranging from initial research questions to ultimate case study conclusions" (Yin, 2003). The chain of evidence for this case study is visualized in Figure 8.

3. *Have key informants review draft case study report.* The final case study tactic to ensure construct validity is to have the draft case study report reviewed by those who have been the subjects of study (Yin, 2003). Two participants, both from the 'Datalogistiek' team (internal data supplier), reviewed the external report.

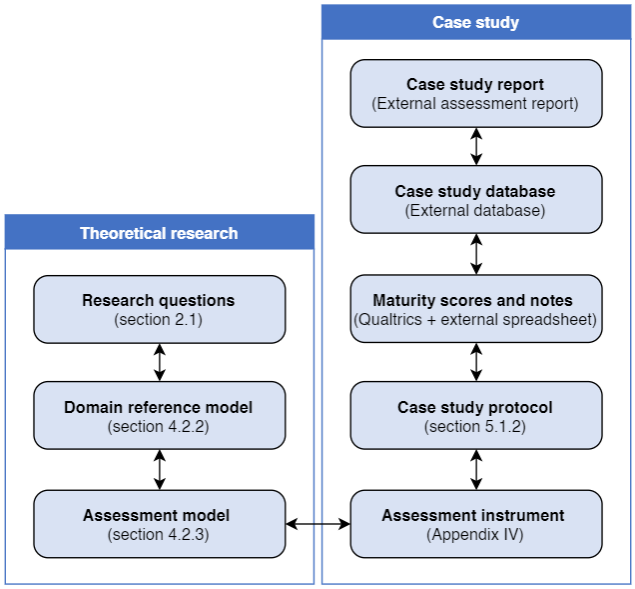


Figure 8 - Chain of evidence (Adapted from Yin, 2003)

Internal validity

As defined by Yin (2003), internal validity is concerned with "establishing a causal relationship, whereby certain conditions are shown to lead to other conditions, as distinguished from spurious relationships". Internal validity testing is only a concern for causal (or explanatory) case studies and not for descriptive/exploratory studies (Yin, 2003), so it is not taking into account for this case study.

External validity

The test on external validity is concerned with "establishing the domain to which a study's findings can be generalized" (Yin, 2003). The maturity model developed in this study is unique and still in an experimental phase. It is also not yet applied in a domain other than the insurance domain, as done in this case study. For this reason, it is not possible to identify the domain to which the study's findings can be generalized. Therefore, external validity is not taken into consideration for this case study. However, the goal is to generalize the model so that organizations in different domains or industries can use it. Future research needs to focus on the multiple application of the DQM3 to achieve this goal.

Reliability

The final case study test, reliability, is concerned with "demonstrating that the operations of a study – such as the data collection procedures – can be repeated, with the same results" (Yin, 2003). The case study tactics to increase the reliability of the case study are to use a case study protocol and develop a case study database:

1. *Use case study protocol*. The protocol used for the case study is described and used throughout this section. As defined earlier, the protocol is constructed using the template from Brereton et al. (2008).

2. *Develop case study database*. The case study database consists of the maturity scores and the notes produced by the respondents. The individual scores of the respondents are automatically stored in Qualtrics. These scores are exported from Qualtrics, stored in a spreadsheet, and used to calculate the overall maturity level at domain sub-component level, domain component level, and domain in general.

5.2 Reflect evolution

The final phase of the development cycle of Mettler (2011) is 'reflect evolution'. In this phase, the design mutability of the DQM3 must be contemplated. According to Mettler (2011), this phase is often neglected, although it is every important: the DQM3 has to be refaced from time to time because the phenomenon under study (DQM) is growing. Mettler (2011) defined the following three decision parameters for this final phase: subject of change, frequency, and structure of change. The decisions made for these parameters are discussed below and summarized in Table 25.

Subject of change

The subject of change is characterized as 'form and functioning'. The form and functioning of the DQM3 is not considered to be unchangeable in the future, as new technologies or future research may provide new best practices for the proactive and continuous management of DQ. Also, extensive research on other maturity models that are not specified to the DQM domain can reveal new insights to change the domain reference model or the assessment model.

Frequency

Because it not clear at this point how often the DQM3 should be revised, the 'frequency' decision parameter is characterized as 'non-recurring'. On one hand, the domain of DQM is complex and dynamic. On the other hand, frequent change will make the DQM3 less usable for organizations, because the company will need to deploy an updated version of the model and perform a new self-assessment. The results of previous assessment cannot then be used by the organization to see progress in maturity.

Structure of change

The structure of change is concerned with the ability of third-parties to change the DQM3. For example, in the case of the CMM model, CMMI is responsible for the development of the maturity model and third-parties are not authorized to change the model. However, the 'structure of change' for the DQM3 is characterized as 'external/open': the opportunity must be given to adapt the maturity model to the context of the organization. Organizations operate in different industries and contexts, so certain sub-components may not apply to a specific business chain and therefore do not have to be assessed by the assessment instrument. However, it is recommended to first use the original assessment instrument to see if the components that do not appear to be 'applicable' in the context of an organization are actually not 'applicable'.

Decision parameter	***Characteristic***			
Subject of change	None	Form	Functioning	**Form and functioning**
Frequency	**Non-recurring**		Continuous	
Structure of change	**External/open**		Internal/exclusive	

Table 25 - Reflect evolution parameters

6. Conclusion, Limitations & Future Research, & Reflection

6.1 Conclusion

How can an organization measure the data quality management maturity level of a business chain?

The main research question defined above was formulated to tackle the problem statement. The following four sub-research questions were formulated to structure the process of answering this question:

SQ1: What are the critical success factors for data quality management?

DQM is "quality-oriented DM, i.e., DM focusing on the collection, organization, storage, processing, and presentation of high-quality data" (Otto et al., 2007). Santos & Lucas (2019) and Lucas (2019) identified Critical Success Factors (CSFs) for (corporate) DQM. These CSFs are essential elements for an organization or project to achieve its mission and should be applied within an organization to ensure a competitive and successful performance (Rockart, 1979; Huang, Wu, & Chen, 2013).

CSFs with similar intents have been merged to come up with a final list of CSFs for DQM. As a result, the following 18 success factors are identified as critical for DQM (in alphabetical order): [CM] Change management; [CDQI] Continuous data quality (management) improvement; [CC] Culture and communication; [DAM] Data architecture management; [DCS] Data customer satisfaction; [DG] Data governance; [DPLM] Data product lifecycle management; [DQAM] Data quality assessment/monitoring; [DQPS] Data quality policies and standards; [DQREM] Data quality requirements management; [DQREM] Data quality risk management; [DSM] Data security management; [DSQM] Data supplier quality management; [D] Documentation; [ET] Education and training; [ECBT] Evaluate cost/benefit tradeoffs; [MCLS] Management commitment, leadership, and support; and [PC] Personnel competency.

SQ2: Which of the existing maturity models allows an organization to perform a self-assessment in which the data quality management maturity level of a business chain is measured?

According to Proença & Borbinha (2016), a maturity model "is a technique that has been proved to be valuable in measuring different aspects of a process or an organization". Seven existing maturity models for the DQM domain were identified and classified with a template based on three dimensions, attributes, defined conditions identified by Mettler et al. (2010) and Van Steenbergen et al. (2007). The four maturity models classified with the 'self-assessment' condition should provide additional supporting materials that can be used to measure the DQM maturity level.

The classification highlighted that only one of the four maturity models has a different condition in the 'support of application' attribute than 'no supporting materials': the DQM maturity model of Kirikoglu (2017), which provides a scorecard. This scorecard also had several limitations, which showed that even this model does not fulfill the requirements of the problem statement. Out of this classification, it could be determined that there is no existing maturity model that allows an organization to perform a self-assessment in which the DQM maturity level of a business chain is measured.

SQ3: How could the identified critical success factors for data quality management be combined in a maturity model to measure the data quality management maturity level of a business chain?

The developed maturity model is called 'Data Quality Management Maturity Model' (DQM3) and consists of two concepts adapted from the CDQM MM of Hüner, Ofner, & Otto (2009): a domain reference model and an assessment model. The domain reference model consists of three layers and is populated with domain components (the CSFs for DQM from the first sub-question) and sub-components (identified via an additional literature study). Based on the developed domain reference model, an assessment model is constructed. This assessment model consists of an assessment instrument (online questionnaire) and five general maturity levels. The assessment instrument is finalized by adding appropriate questions and measurement criteria to make it fit for a pretest and case study at Achmea.

SQ4: How is the designed maturity model used in practice?

Two participants experienced no technical difficulties while completing the questionnaire during a pretest. Both participants also reacted positively to the components (and questions) included in the assessment instrument. They also confirmed that those indeed influence DQM and recognized the importance of these components. As a result of this pretest, only a few improvements have been made to the formulation of some questions.

With the improved DQM3, a case study was performed in which a total of 128 respondents (108 internal data suppliers, 2 external data suppliers, and 18 data customers) from 'the data logistic chain (Datalogistiek)' of Achmea participated. Overall, the DQM3 performed well in practice. Just as with the pretest, only a few improvements were made to the formulation of some questions. These questions were changed to avoid interpretation problems in future maturity assessment.

The case study also showed that the space for comments at the end of each domain component block (added to provide Achmea with descriptive data outside the boundaries of the theoretical design) was beneficial in identifying potential improvement areas. This data is used by an organization to plan specific improvements.

RQ: How can an organization measure the data quality management maturity level of a business chain?

The DQM maturity level of a business chain can be measured with the DQM3 developed in this study. The DQM3 consists of a domain reference model and an assessment model, and can be applied via three steps. First, an organization should identify the teams/departments involved in the business chain and place them under the right contingency group. Second, the assessment instrument should be distributed to all individuals that are part of the teams/departments via email or with an anonymous survey link. Finally, the results should be analyzed, an overall maturity level should be assigned (at domain, domain component, and sub-component level), and improvements should be planned.

Despite the promising results that the DQM3 performs well in practice, as described above, the maturity model still has several limitations and thus require more in-depth research. The limitations and possibilities for future research are discussed in the next section.

6.2 Limitations and future research

The first limitation of this study regards the domain sub-components included in the domain reference model. The publications of Santos & Lucas (2019) and Lucas (2019) and the sources they used for their studies were consulted to identify these sub-components. However, some of the publications used by Santos & Lucas (2019) were not available in one of the seven search engines selected for this study. As a result, these publications could not be consulted. Some sub-components that may affect a specific domain component may, therefore, not have been identified. However, all current sub-components included in the domain reference model are supported by at least two references. I am therefore convinced that the sources used to populate the domain reference model provided sufficient input to identify all domain sub-components that have a significant impact on DQM maturity.

Another limitation regarding these components is the exclusion of the CSF 'data quality policies and standards' as domain component in the domain reference models. As discussed in this thesis report, the only relevant information that could be extracted from the sources that are used by Lucas (2019) and Santos & Lucas (2019) was that according to Xu (2015) and Xu, Koronios, & Brown (2003), an organization should have in place "appropriate (simple, relevant, and consistent) DQ policies and standards" by establishing and implementing/enforcing those policies and standards. However, it was not defined what exactly is meant by 'policies and standards'. Therefore, the role of this CSF should be further explored in future research.

A third limitation regarding these domain sub-components and the domain components is the order and grouping of these components in the domain reference model. The logical sequence and grouping of those domain components are based on two studies: the CSF framework created by Baškarada & Koronios (2014) and the total quality management implementation three-level framework developed by Hietschold, Reinhardt, & Gurtner (2014). This study does not test whether these components fit the three defined parts of the domain reference model. Future research could include a factor analysis (as done in the study of Xu, 2013) to test the parts of the domain reference model for their validity.

The fourth limitation of this study is that only two employees of the same organization (Achmea) participated in the pretest to validate the domain sub-components included in the domain reference model. These two employees were asked during an online session to provide feedback on whether the sub-components contribute to DQM maturity. Although both participants responded positively on the domain sub-components, confirmed that those indeed influence DQM, and recognized the importance of these components, some bias may occur.

The fifth and final limitation of this study regards the research design used in this study: the design cycle of Wieringa (2014). This cycle limits to the validation of the DQM3 in practice, whereas the engineering cycle of Wieringa (2014) should be used to transfer the validated treatment into the real world and evaluate whether this treatment was successful. This is related to the examination of a case study in this study: 'the data logistic chain (Datalogistiek)' at Achmea. However, due to time constraints and limited resources, only one business chain at a large organization (> $100 million, categorization defined by Xu (2003)) has been examined for now. Future research should focus on the multiple application of the DQM3 in other sizes of organizations (very-small, small, and medium-sized organization) and industries (besides the insurance industry) to identify the domain to which the findings of this study can be generalized (i.e., establish external validity).

6.3 Reflection on research goals and contributions

As discussed in the introduction of this thesis, the goal of this research was twofold. The first goal was to tackle the problem statement by developing a DQM maturity model that allows an organization to perform a self-assessment and provides the ability to measure the DQM maturity level of a business chain. With the DQM3, this research goal is achieved. This thesis report will, therefore, contribute to the scientific body of knowledge and serve as a foundation for future research, as discussed in the previous subsection.

The second goal of this research was to apply and validate the designed DQM3 by performing a case study at Achmea. To apply and validate the DQM3, not only a case study was conducted: the pre-test also contributed to this. However, this research goal is partially achieved since the DQM3 is not validated on its external validity. The results of the case study, on the other hand, are used by Achmea to improve the DQM maturity in 'the data logistic chain (Datalogistiek)', making the practical contribution of this study.

6.4 Personal reflection

With the completion of this thesis project, I would like to briefly reflect on the 'intended learning outcomes' of this study. The first phase of this thesis project was about defining the research method for this thesis and conducting a scientific survey of the literature in the field of study. With the knowledge and skills I gained during my master Business Informatics, I was able to design and develop such a research plan and demonstrate a thorough understanding of the literature on DQM and maturity models based on a defined literature research protocol. During this first phase, I learned to deal with definitions and existing models to get the best possible picture of the current state of literature.

The second phase was concerned with the actual execution of the research according to the approach defined in the first phase, i.e., conduct sound scientific research according to a predefined plan. With the completion of this thesis report, I learned how to perform such a large thesis project. I also learned how to write a scientific report about the conducted research. Although I was a little behind schedule because of the extensive literature study to develop the maturity model, I am happy with the result that I contribute to the scientific body of knowledge, as discussed in the previous section. With a few tips from my project supervisor Nico and first supervisor Ari, I also managed to increase the readability of this thesis report.

References

Akhavan, P., Jafari, M., & Fathian, M. (2006). Critical Success Factors of Knowledge Management Systems: a Multi-Case Analysis. *European Business Review Journal, 18*(2), 97-113.

Ariyachandra, T. R., & Folick, M. N. (2008). Critical Success Factors in Business Performance Management—Striving for Success. *Information Systems Management, 225*(2), 113-120. doi:10.1080/10580530801941504

Bacon, C. J., & Fitzgerald, B. (2001, April). A systematic framework for the field of information systems. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 32*(2), 46-67. doi:10.1145/506732.506738

Baškarada, S., & Koronios, A. (2014). A critical success factor framework for information quality management. *Information Systems Management, 31*(4), 276-295. doi:10.1080/10580530.2014.958023

Baškarada, S., Koronios, A., & Gao, J. (2007). IQM-CMM: a framework for assessing organizational information quality management capability maturity. *12th International Conference on Information Quality (ICIQ07).* Massachusetts Institute of Technology: Cambridge, MA, USA. doi:10.2139/ssrn.2142971

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009). Developing Maturity Models for IT Management. *Bus. Inf. Syst. Eng, 1*, 213-222. doi:10.1007/s12599-009-0044-5

Black, S. A., & Porter, L. J. (1996). Identification of the critical factors of TQM. *Decision Sciences, 27*(1), pp. 1-21. doi:10.1111/j.1540-5915.1996.tb00841.x

Brereton, P., Kitchenham, B. A., Budgen, D., & Li, Z. (2008, June). Using a Protocol Template for Case Study Planning. *12th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, (pp. 41-48). doi:10.14236/ewic/EASE2008.5

Caballero, I. (2008). IQM3: Information Quality Management Maturity Model. *Journal of Universal Computer Science, 14*(22), 3658-3685.

Chrissis, M. B., Konrad, M., & Shrum, S. (2003). *CMMI guidelines for process integration and product improvement.* Boston: Addison-Wesley.

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

Davenport, T. (2006). Competing on Analytics. *Harvard Business Review, 84*(1), 98-107.

Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what they know.* Harvard Business Press.

de Bruin, T., Rosemann, M., Freeze, R., & Kulkarni, U. (2005). Understanding the Main Phases of Developing a Maturity Assessment Model. *Australasian Conference on Information Systems (ACIS).* Australia, New South Wales, Sydney.

Donaldson, L. (2001). *The contingency theory of organizations.* Sage.

Eckerson, W. W. (2002). Data Warehousing Special Report: Data quality and the bottom line. *Applications Development Trends*.

English, L. P. (1999). *Improving data warehouse and business information quality: methods for reducing costs and increasing profits* (1 ed.). New York: Wiley.

Fraser, P., Moultrie, J., & Gregory, M. (2002). The use of maturity models/grids as a tool in assessing product development capability. *IEEE international engineering management conference*, *1*, pp. 244-249. Cambridge, UK. doi:10.1109/IEMC.2002.1038431

Gericke, A., Rohner, P., & Winter, R. (2006). Networkability in the Health Care Sector - Necessity, Measurement and Systematic Development as the Prerequisites for Increasing the Operational Efficiency of Administrative Processes. *ACIS 2006 Proceedings.*

Gillies, A., & Howard, J. (2003). Managing change in process and people: combining a maturity model with a competency-based approach. *Total Quality Management & Business Excellence, 14*(7), 779-787. doi:10.1080/1478336032000090996

Haegemans, T. (2018). *Essays on data quality management with applications in the financial industry.*

Harexian, A. (2019). *The impact of poor data quality: A case study in an insurance company.*

Haug, A., Zachariassen, F., & van Liemph, D. (2011). The cost of poor data quality. *Journal of Industrial Engineering and Management, 2*, 168-193. doi:10.3926/jiem.2011.v4n2.p168-193

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004, March). Design Science in Information Systems Research. *MIS Quarterly, 28*(1), 75-105. doi:10.2307/25148625

Hietschold, N., Reinhardt, R., & Gurtner, S. (2014). Measuring critical success factors of TQM implementation successfully - a systematic literature review. *International Journal of Production Research, 52*(21), 6252-6272. doi:10.1080/00207543.2014.918288

Huang, S. J., Wu, M. S., & Chen, L. W. (2013). Critical sucess factors in aligning IT and business objectives: A Delphi study. *Total Quality Management & Business Excellence, 24*(9-10), 1219-1240. doi:10.1080/14783363.2011.637785

Hüner, K. M., Ofner, M., & Otto, B. (2009). Towards a maturity model for corporate data quality management. *Proceedings of the ACM symposium on Applied Computing*, (pp. 231-238). doi:10.1145/1529282.1529334

Jayawardene, V., Sadiq, S., & Indulska, M. (2015). *An Analysis of Data Quality Dimensions.*

Kirikoglu, O. (2017). *A maturity model for improving data quality management.* University of Twente.

Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews.* Keele, UK: Keele University.

Loshin, D. (2011). *The Practitioner Guide to Data Quality Improvement.* USA: Elsevier.

Lucas, A. (2010). Corporate data quality management: From theory to practice. *5th Iberian Conference on Information Systems and Technologies* (pp. 1-7). Santiago de Compostela: IEEE.

Lucas, A. (2019). Critical Success Factors for Corporate Data Quality Management. In Á. Rocha, H. Adeli, L. P. Reis, & S. Costanzo, *New Knowledge in Information Systems and Technologies* (pp. 630-644). Springer, Cham. doi:10.1007/978-3-030-16181-1_60

Mettler, T. (2011, November). Maturity assessment models: a design science research approach. *International Journal of Society Systems Science (IJSSS), 3*(1/2), 81-98. doi:10.1504/IJSSS.2011.038934

Mettler, T., & Rohner, P. (2009). Situational maturity models as instrumental artifacts for organizational design. *Proceedings of the 4th international conference on design science research in information systems and technology.* New York, NY, USA: Association for Computing Machinery. doi:10.1145/1555619.1555649

Mettler, T., Rohner, P., & Winter, R. (2010). Towards a Classification of Maturity Models in Information Systems. In A. D'Atri, M. De Marco, A. Braccini, & F. Cabiddu, *Management of the Interconnected World* (pp. 333-340). Physica-Verlag HD. doi:10.1007/978-3-7908-2404-9_39

Mosley, M. (2008). *The DAMA Dictionary of Data Management.* LLC.

Ofner, M., Otto, B., & Österle, H. (2013). A Maturity Model for Enterprise Data Quality. *Enterprise Modelling and Information Systems Architectures - An International Journal, 8*(2), 4-24. doi:10.18417/emisa.8.2.1

Otto, B., Wende, K., Schmidt, A., & Osl, P. (2007). Towards a Framework for Corporate Data Quality Management. *18th Australasian Conference on Information Systems* (pp. 916-926). Toowoomba, Australia: The University of Southern Queensland.

Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. V. (1993). Capability maturity model, version 1.1. *IEEE Software, 10*(4), 18-27.

Porter, L. J., & Parker, A. J. (1993). Total quality management - the critical success factors. *Total quality management, 4*(1), pp. 13-33.

Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM, 41*(2), 79-82.

Rockart, J. F. (1979). Chief executives define their own data needs. *Harbard business review*, 81-93.

Rosemann, M., & De Bruin, T. (2005). Towards a Business Process Management Maturity Model. *ECIS 2005 Proceedings of the Thirteenth European Conference on Information Systems* (pp. 1-12). Regensburg: Verlag and the London School of Economics.

Ryu, K. S., Park, J. S., & H., P. J. (2006). A Data Quality Management Maturity Model. *ETRI Journal, 28*(2), 191-204. doi:10.4218/etrij.06.0105.0026

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), 433-438. doi:10.5281/zenodo.3455599

Saraph, J., Benson, P., & Schroeder, R. (1989). An instrument for measuring the critical success factors of quality management. *Decision sciences, 20*(4), pp. 810-829.

Strong, D., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. (110, Ed.) *Communications of the ACM, 40*(5), 103.

Sue, V. M., & Ritter, L. A. (2007). *Conducting Online Surveys.* Thousands Oaks, California: Sage Publications, Inc.

Tayi, G. K., & Ballou, D. P. (1998). Examing Data Quality. *Communications of the ACM, 41*(2), 54-57.

Tee, S. W., Bowen, P. L., Doyle, P., & Rohde, F. H. (2007). Factors influencing organizations to improve data quality in their information systems. *Accounting & Finance, 47*(2), 335-355.

Vagias, W. M. (2006). *Likert-type scale response anchors.* Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University.

van Steenbergen, M., van den Berg, M., & Brinkkemper, S. (2008). A Balanced Approach to Developing the Enterprise Architecture Practice. In J. Filipe, J. Cordeiro, & J. Cardoso, *Enterprise Information Systems* (Vol. 12, pp. 240-253). Springer, Berlin, Heidelberg. doi:10.1007/978-3-540-88710-2_19

Weber, K., Otto, B., & Österle, H. (2008). One size does not fill all - a contingency approach to data governance. *Journal of Data and Information Quality (JDIQ), 1*(1).

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*(2), xiii-xxiii.

Wieringa, R. J. (2014). *Design science methodology for information systems and software engineering.* Springer-Verlag Berlin Heidelberg.

Wixom, B. H., & Watson, H. J. (2001). An empirical investigation of the factors affecting data warehousing success. *MIS Quarterly, 25*(1), 17-41.

Xu, H. (2003). Would organzation size matter for data quality. *Proceedings of the Eights International Conference on Information Quality*, (pp. 365-379).

Xu, H. (2013). *Factor analysis of critical success factors for data quality.*

Xu, H. (2015). What are the most important factors for accounting information quality and their impact on AIS data quality outcomes? *ACM J. Data Inf. Qual., 5*(4). doi:10.1145/2700833

Xu, H., & Lu, D. (2003). The critical success factors for data quality in accounting information system - different industries' perspective. *Issues Inf. Syst, 4*, 762-768.

Xu, H., Koronios, A., & Brown, N. (2003). Managing data quality in accounting information systems. In L. Z. Joia, *IT-Based Management: Challenges and Solutions* (pp. 277-299). doi:10.4018/978-1-59140-033-2.ch016

Yeoh, W., & Koronios, A. (2010). Critical success factors for business intelligence systems. *Journal of computer information systems*(50), 23-32.

Yin, R. K. (2003). *Case study research: Design and Methods.* Thousands Oaks: Sage Publications.

Zhang, L., Lee, M. K., Zhang, Z., & Banerjee, P. (2003). Critical Success Factors of Enterprise Resource Planning Systems Implementation Success in China. *36th Hawaii International Conference on System Sciences*, (pp. 1-10).

Appendix

Appendix I – Interview about the DMM model of the CMMI Institute within Achmea (Dutch)

Geïnterviewde: Data Architect binnen Achmea

Datum: 26-11-2019

Korte introductie van DMM

Binnen Achmea willen we een data gedreven organisatie worden en een aantal vragen staan hierin centraal: wat is dat nou precies, waar staan we nu eigenlijk en waar willen we naar toe? Daar heb je een methodologie voor nodig en binnen Achmea is gekozen om het DMM model te gebruiken. Binnen Achmea wordt ook het CMMi voor processen gebruikt en deze sluit goed aan bij het DMM model. Daarnaast is DMM heel compleet vergeleken met DAMA-BOK.

1. Kun je in het kort uitleggen wat het DMM model doet voor Achmea en waarom deze door Achmea gekozen is om de maturity te testen?

DMM is een raamwerk. Binnen dit raamwerk zijn er zes domeinen gedefinieerd: data management strategy, data governance, data quality, data operations, platform & architecture, en supporting processes. Bij elk van deze gebieden binnen Achmea is een 'area tracker' aangewezen. Waar ik zelf verantwoordelijk ben zijn 'data quality' en 'data operations'. Elk van deze zes domeinen bestaat uit een aantal sub domeinen om het behapbaar te maken: bijv. data quality bestaat uit data quality strategy, data profiling, data quality assessment, en data cleansing.

Ook zijn er vijf verschillende niveaus gedefinieerd binnen DMM. Binnen Achmea hebben we hier een level aan toegevoegd: niveau 0, zijnde "we doen helemaal niks". Officieel bestaat het niveau niet binnen DMM: binnen DMM is het laagste niveau, niveau 1: we voeren processen ad-hoc uit.

2. Hoe wordt dit model binnen Achmea getest? Self-assessment met een vragenlijst? Wordt dit per afdeling gecheckt (welke scope)?

In 2015 hebben we een eerste meting laten uitvoeren door een externe partij door middel van interviews. Dit is gedaan bij FBTO maar staat ook karakteristiek voor heel Achmea. Hier scoorde Achmea op heel veel plekken tussen de 0 en 1. Vanuit daar is gezegd om stap voor stap omhoog te gaan: in 2017-2018 willen we op level 2 zitten en eind 2018-2019 willen we op level 3 zitten.

Inmiddels is binnen Achmea een quickscan (in Excel) ontwikkeld waar je op basis van dezelfde aantal vragen (25) kunt zien waar je op staat. Een vragenlijst is gemaakt waarin voor elk sub domein een basisvraag is opgesteld. Deze basisvraag heeft vijf mogelijke antwoorden. Afhankelijk van het antwoord waar je als organisatie, keten of afdeling het meest aan voldoet, kun je zien op welk niveau je zit. Uit deze vragenlijst komt automatisch een spinnenweb uitrollen.

De quickscan is toepasbaar op een bepaalde keten, afdeling of team. Deze quickscan kan gemaakt worden door een senior manager om te zien welk niveau hun team of het stuk waar ze aan werken zitten. Op basis hiervan kan gekeken worden waarop ze slecht scoren zodat besloten kan worden binnen welk DMM sub domein men verbetering wil aanbrengen.

3. Hoe worden de resultaten gecheckt of het daadwerkelijk ook zo is dat een team of afdeling ook echt het level scoort welke ze aangegeven hebben?

Om dit te voorkomen is een 'dataset scan' gemaakt en dan gaan we kijken naar een bepaalde dataset of daar gevoelige data in staat. Als dit gevonden wordt dan worden onder andere de volgende vragen gesteld: wie heeft hem daar neergezet, hoe komt het daar, wie heeft er toegang toe, etc.? Daar is een vragenlijst voor opgesteld en de antwoorden van deze vragenlijst moeten corresponderen met de resultaten uit de quickscan. Als deze twee niet corresponderen weten we dat er politiek correcte antwoorden gegeven zijn.

4. Hoe vaak wordt / is het model getest? Wat is de gewenste frequentie?

Aangezien de quickscan ongeveer twee maanden geleden ontwikkeld is, wordt deze op dit moment nog niet overal toegepast. De gewenste frequentie is 1x per jaar. We willen ook de vorige weten en zien dat er progressie is tussen de antwoorden.

5. Is het bekend welke problemen er binnen de datalogistiek afdeling spelen?

De grootste problemen binnen de datalogistiek afdeling is het feit dat ze onvoldoende in kaart hebben gebracht welke informatie ze eigenlijk binnen krijgen, welke data naar de Rabobank gestuurd wordt en aan welke kwaliteitseisen deze data moet voldoen. Daar wordt nauwelijks aandacht aan besteed met als gevolg dat er veel operationele problemen zitten binnen de datalogistiek.

6. Wat mist er volgens jou aan dit model? Is er bijv. te weinig focus op data kwaliteit?

Het enigste wat op dit moment mist is de implementatie binnen de verschillende afdelingen. De quickscan is afgelopen jaar ontwikkeld en sinds twee maanden is het functioneel. De gereedschapskist ligt er, de opzet en het bestaan is er. Alleen de werking is minder omdat niet iedereen er evenveel "zin" in heeft.

Appendix II – Critical Success Factors for (Corporate) Data Quality Management

A: CSFs for DQM identified by Santos & Lucas (2019)

CSF	Description
1. DQ policies and standards	• Implementation of a standard methodology. • Standardizing codes, rules, and definitions. • Reformulation of the data model, when necessary.
2. Input controls	• Quality assurance of data loading. • Certification of external data sources. • Certification of existing data. • Implementation of robust validation routines in data collection. • Controlling the way internal data is generated.
3. Production of a strategic plan for DQ	• Strategic alignment with the business. • Knowledge of maturity level of the organization. • DQ vision articulation with business.
4. Organizational culture with a focus on DQ	• Perception of the importance of DQ. • Coherence in the way DQM processes are applied in the organization. • Involvement with the academic environment. • Focus on consumers. • Market and law analysis. • Compliance with contractual obligations. • The quality of the data should be seen as a critical issue of the business and dealt continuously and proactively. • Development of a culture of motivation, trust, and respect.
5. Top management commitment and support	• Recognition by the top management of the importance of DQ. • Authorization to support activities related to DQ and attribution of rewards to employees. • Change in the work environment to enable employees to accept the importance of DQ. • Implementation of measures aimed at motivating people in the organization to support the DQ initiative and the inherent organization changes. • Obtaining necessary operational resources such as financing and personal competences.
6. Data governance	• Set of essential actions to ensure data compliance with organization strategies.
7. Continuous improvement	• Institutionalization of continuous improvement of DQ. • Identification and troubleshooting. • Perform data cleansing. • Monitor progress towards DQ objectives by holding periodic presentations and communications meetings. • Projects to improve the quality of data should be part of the company's budget.
8. Internal and external monitoring and evaluation	• Implementation of ETL and DQ tools. • Identification of problems such as missing data, incorrect values, duplicates records, and violations of business rules. • Establishment of service level agreements. • Implementation of results measurement. • Performance evaluation. • Implementation of benchmarking techniques. • Implementation of statistical process control. • Definition of metrics.
9. Change Management	• Change in organizational processes and behaviors. • Adapting the rules of data integrity considering the changes in business processes and requirements. • Existence of organizational competencies to manage internal and external changes. • Reengineering and process integration. • Change of culture at all levels of the organization.
10. Conducting regular audits	• Identification of problems such as missing data, incorrect values, duplicate records, and violations of business rules.

	• Identification of opportunities, deficiencies, and gaps. • Ensuring that the appropriate controls are in place.
11. Architecture Management	• Appropriate software and hardware acquisitions. • Update applications. • Adequate technological capacities. • Minimization of interfaces. • Data integration. • Implementation of data warehouse.
12. User Focus	• User engagement. • Focus on users' needs and quality requirements. • Active participation of users in order to ensure and improve DQ.
13. Education and Training	• Acquisition of new competences covering the entire organization, from the top and intermediate management to the collaborators. • Providing effective and appropriate initial and continuous training to employees.
14. Appointment of managers and definition of roles	• Definition of responsibilities for DQ. • Identification of owners and the custodians. • Appointment of data stewards and a data champion. • Appointment of a specialist or a group of experts as managers of DQ.
15. Documentation	• Elaboration of adequate and sufficient documentation, both at the user and the data administrator levels. • Documentation of all data items.
16. Communication	• Sharing knowledge and communication between different departments, within departments, and among different professionals. • Sharing between employees needs a strong, reliable culture and also transparency throughout the organization.
17. Middle management commitment and support	• Acceptance of responsibility for DQ performance by middle managers. • Effective procedures at middle management level.
18. Teamwork	• Definition of a team for DQ. • Centralization of competencies, both technical and the interpersonal, in a team of excellence.
19. Security and internal control	• Access control and permissions. • Implementation of appropriate internal controls to systems and processes, including security control. • Analysis of logs of user activities. • Control of data privacy violations. • Controls at people level, such as segregation of functions.
20. Risk Management	• Identification, analysis, monitoring, prioritization and categorization of risks. • Implementation of risk mitigation procedures.
21. Sufficient resources	• Allocation of sufficient resources: technical, monetary, people, competencies, and time.
22. Storage Management	• Policies of backup and retention. • Implementation of a repository of meta data. • Selection, preservation, and management of digital data to facilitate and future discovery and recovery of this data. • Implementation of reuse practices and data preservation.
23. Evaluate cost/benefit tradeoffs	• Tracking costs and benefits and identifying the critical point where a more centralized business initiative is justified. • Identification of costs caused by poor DQ. • Have systematic cost/benefit analysis of DQ controls and activities in order to maximize benefits at minimum cost.

Table 26 - Critical success factors for data quality management (Adapted from Santos & Lucas, 2019)

B: CSFs for DQM identified by Lucas (2019)

CSF	Description
1. Data Governance	This includes a set of key actions to ensure data compliance with organizational strategies. It defines a suitable organizational structure to produce high quality information. It should define responsibilities for the DQ: identify the owners and custodians; appoint data stewards and a data champion: appoint an expert or a group of experts as DQ managers (da C. dos Santos, 2015). It should promote teamwork between business and IT people, as a key to improving data quality.
2. Management Commitment and Leadership	Top management must form a sound foundation for clear values and data quality policies and provide the corresponding resources. Companies must integrate data quality into the organizational strategy to achieve consistent and lasting excellence (Saraph, Benson, & Schroeder, 1989).
3. Continuous Data Quality Management Improvement	Continuous DQM improvement deals with using Key Performance Indicators (DQM-KPI) to continuously monitor the effectiveness of organizational DQM efforts (Baškarada & Koronios, 2014). 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 validation (Xu, Koronios, & Brown, 2003).
4. Data Architecture Management	Architecture of the IS ecosystem or its geography is relevant to the type of DQ initiative. 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.
5. Culture and Communication	This involves encouragement of an organization-wide culture committed to data quality improvement (Black & Porter, 1996). Communication is viewed as a two-way process with feedback channels available. Communication is seen as an ongoing process, taking into account ways of strengthening concepts in the future (Porter & Parker, 1993).
6. Data Quality Policies and Standards	The organization should have data quality policies and standards that are simple, relevant and consistent. There are two main components: 1. Establishing appropriate and specific data quality policies and standards; 2. Implementing/enforcing policies and standards.
7. Data Quality Assessment/Monitoring	Before any DQ improvement can be attempted, the current state of DQ first needs to be assessed both qualitatively and quantitatively (Baškarada & Koronios, 2014). Profiling tools can be used to access most of the data quality dimensions. Qualitative and quantitative DQM metrics or Key Performance Indicators (DQ-KPIs) should be defined, and then used to continuously monitor the effectiveness of organizational DQM efforts (Baškarada & Koronios, 2014). At certain time intervals DQ should be monitored using the same data profiling tools. In addition, compliance with policies and standards should be monitored.
8. Data Quality Requirements Management	It is important to identify all the key stakeholders and collect and model their requirements (Baškarada & Koronios, 2014).
9. Focus on Data Customer Satisfaction	This entails focusing on data customers' needs and their quality requirements. It should enable active participation from data customers to ensure and improve data quality (Xu, Koronios, & Brown, 2003). "Data customers" can refer to the client (external customers) and the internal customer.
10. 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) (Baškarada & Koronios, 2014). Clarity of process ownership (process owners), boundaries, and steps must be established (Saraph, Benson, & Schroeder, 1989).
11. Personnel Competency	The competence of personnel responsible for DQ is particularly important. For instance, employees should be exceptionally skilled and informed in both technical and business areas (Xu, Koronios, & Brown, 2003).
12. Management of Changes	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 change include things such as government regulations, technology, economy, and market changes (Xu, Koronios, & Brown, 2003).

13. Data Security Management	Access security is a key DQ dimension and data security management requires an organization to have effective access controls in place. The controls must ensure that all users are appropriately authenticated as well as authorized with the least set of privileges they require. For instance, IS developers should not have access to the production environment. Furthermore, audit trails (logs of user' activities on the IS) should be analyzed (e.g., for exceptions) and periodically reviewed (Baškarada & Koronios, 2014).
14. Data Quality Risk Management	Risk management can be defined as the awareness of and the level of commitment to the reduction of the consequences of poor DQ (Xu, Koronios, & Brown, 2003). DQ risks to business objectives (including financial risks, reputation risks, regulatory risks, etc.) should be diagnosed, documented, analyzed, classified, prioritized and mitigated/controlled. Effective DQ Risk Management should allow organizations to focus their DQM efforts on the most critical information products, thus, increasing DQM efficiency and effectiveness (Baškarada & Koronios, 2014).
15. Audit and Reviews	It is important to have independent internal and external regular data quality audits and reviews to ensure appropriate controls are in place (Xu, Koronios, & Brown, 2003) .
16. Training	Training in data quality concepts, methods and tools is a precondition for employee involvement and empowerment (Saraph, Benson, & Schroeder, 1989). Relevant training needs should be identified and documented, and training workshops should be conducted regularly. In additional to formal training, mentoring programs should ensure on-the-job professional development (Baškarada & Koronios, 2014).
17. Understanding of the IS and the relevance of Data Quality	It is important to understand how the information systems work (technical competence) and IT personnel and data consumers need to understand the importance of data quality (Xu, Koronios, & Brown, 2003) .
18. Evaluate Cost/Benefit Trade-offs	Before any process improvements are made, it is critical to estimate the costs associated with poor DQ and corresponding improvement initiatives, as well as many potential or cost savings that may result from any process improvements (Baškarada & Koronios, 2014).
19. Supplier Partnership	"Data supplier quality management means to have an effective data quality management relationship with raw data suppliers, which has two important parts: 1. To have agreement about the acceptable level of quality of raw data to be supplied, such as the requirements of availability, timeliness, accuracy and completeness; 2. To provide regular data quality reports and technical assistance to data suppliers" (Xu & Lu, 2003) p.291

Table 27 - Critical success factors for corporate data quality management (Adapted from Lucas, 2019)

Appendix III – Blueprint of the assessment instrument (version I)

Welcome!

Thank you for taking the time to participate in this survey to factors influencing data quality management within:

'[*the business chain*]'

[ADD EXPLANATION ON SCOPE OF BUSINESS CHAIN]

Participation in this survey allows you to give your opinion about your experiences with critical success factors that influence data quality management within the chain. The data you leave will be used to make improvement within the chain. You are completely anonymous while completing this survey.

Contingency questions

Logic	Code	Questions and answer options
	[CQ-1]	What is your role within [*the business chain*]? Internal data supplier | External data supplier | Data customer
If [CQ-1] = Internal data supplier	[TD-a]	Select one or more teams/departments you are part of below: *[Team/department name 1]* | *[Team/department name 4]* | *[Team/department name 7]* *[Team/department name 2]* | *[Team/department name 5]* | *[Team/department name 8]* *[Team/department name 3]* | *[Team/department name 6]* | *[Team/department name N]*
If [CQ-1] = External data supplier	[TD-b]	Select one or more teams/departments you are part of below: *[Team/department name 1]* | *[Team/department name 2]* | *[Team/department name N]*
If [CQ-1] = Data customer	[TD-c]	Select one or more teams/departments you are part of below: *[Team/department name 1]* | *[Team/department name 3]* | *[Team/department name 5]* *[Team/department name 2]* | *[Team/department name 4]* | *[Team/department name N]*

Questions that assesses the domain reference model

Block 1. [MCLS] Management commitment, leadership, and support

For the questions below, *management* is not confined to top or senior management, but refers to all levels of management within [*the organization*].

If [CQ-1] = Internal data supplier

[MCLS-1] Recognition of importance and responsibility

To what extent do you agree with the following statements?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Management [*within the organization*] recognizes the importance of data quality.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Management [*within the organization*] takes responsibility for the quality of the data.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

[MCLS-2] Allocation of resources

To what extent does management allocates sufficient resources to support data quality within [*your team(s)/department(s)*]?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Sufficient **budget**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Sufficient **technical tools (software)**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Sufficient **expertise**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Sufficient **skilled personnel**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Sufficient **time**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

[MCLS-3] Attribution of rewards

How often does management attributes rewards to employees within [*your team(s)/department(s)*] who are proactively committed to managing data quality?

- Examples of rewards: recognition for data quality improvement suggestions; increased budget; positive feedback.

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Never	Rarely	Sometimes	Often	Very often

[MCLS-4] Strategic data quality plan

To what extent do you agree with the following statement?
There is a clear plan for improving data quality within the [*the business chain*].

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Strongly disagree	Disagree	Neutral	Agree	Strongly agree

Block 2. [DG] Data governance

If [CQ-1] = Internal data supplier

[DG-1] Organizational structure

To what extent do you agree with the following statement?
The organizational structure within [*the organization*] is suitable to produce high-quality data and there are clear roles and responsibilities regarding data quality?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Strongly disagree	Disagree	Neutral	Agree	Strongly agree

[DG-2] Teamwork

To what extent is there effective teamwork between business and IT people (within and between different teams/departments)?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Teamwork **within** [*your team(s)/department(s)*]	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective
Teamwork **between** [*your team(s)/department(s)*] and other teams/departments within [*the business chain*]	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective

Block 3. [CC] Culture and communication

If [CQ-1] = Internal data supplier

[CC-1] Organizational culture focused on data quality

To what extent does [*the organization*] have a culture that focuses on the continuous and proactive management of data quality?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused

[CC-2] Communication

To what extent is there effective communication between business and IT people (within and between different teams/departments)?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Communication **within** [*your team(s)/department(s)*]	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective
Communication **between** [*your team(s)/department(s)*] and other teams/departments within [*the business chain*]	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective

	[CC-3] Knowledge sharing	How often is knowledge shared within [*the organization*] at knowledge sharing occasions and via networks of experts?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Regular knowledge sharing (repeated at specific intervals, e.g. once every month)	Never	Rarely	Sometimes	Often	Very often
Event-triggered knowledge sharing (at specific events, e.g., project's ends, coming up of a new technology)	Never	Rarely	Sometimes	Often	Very often
Network of experts (e.g., knowledge teams or centers, communities of practice)	Never	Rarely	Sometimes	Often	Very often

Block 4. [ET] Education and training

If [CQ-1] = Internal data supplier

[ET-1] Training workshops/ programs

To what extent does [*the organization*] offer sufficient training on data quality management?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Sufficient **initial training** on data quality management (new personnel, new/ upgrade system).	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Sufficient **ongoing training (apart from the initial training)** on data quality management offered (regular training for employees/managed, e.g. once every year).	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

[ET-2] Professional development

To what extent is on-the-job professional development of employees supported by [*the organization*] (e.g., by providing mentoring programs)?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not supported at all	Slightly supported	Somewhat supported	Very supported	Extremely supported

Block 5. [DQREM] Data quality requirements management

If [CQ-1] = Internal data supplier

[DQREM-1a] Stakeholder management

To what extent is data quality central to prioritizing the requests that [*your team(s)/department(s)*] realizes?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not central at all	Slightly central	Somewhat central	Very central	Extremely central

[DQREM-2] Data (quality) requirements management

To what extent do you agree with the following statements?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Data requirements are accurately captured within [*your team(s) /departments*] by modeling them in conceptual, logical, and physical data models.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Data quality requirements are proactively identified, verified, validated, and updated within [*your team(s)/departments*] in consultation with stakeholders.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

If [CQ-1] = Data customers

[DQREM-1b] Stakeholder management

To what extent is [*the business chain*] focused on the needs and quality requirements that [*your team(s)/ department(s)*] places on the supplied data?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Needs of [*your team(s)/ department(s)*]	Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused
Quality requirements that [*your team(s)/ department(s)*] places on the supplied data.	Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused

Block 6. [DQRIM] Data quality risk management

If [CQ-1] = Internal data supplier

[DQRIM-1] Awareness of consequences

To what extent is [*the organization*] aware of the consequences that poor data quality within [*the business chain*] can have?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not aware at all	Slightly aware	Somewhat aware	Very aware	Extremely aware

[DQRIM-2] Commitment to impact reducing

To what extent is [*the organization*] committed to reducing the impact that poor data quality within [*the business chain*] can have?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not committed at all	Slightly committed	Somewhat committed	Very committed	Extremely committed

Block 7. [DQAM] Data quality assessment/monitoring

If [CQ-1] = Internal data supplier

[DQAM-1] Data profiling

To what extent does [*your team(s)/departments*] make adequate use of data profiling to proactively identify data quality problems?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not adequate at all	Slightly adequate	Somewhat adequate	Very adequate	Extremely adequate

[DQAM-2] Data quality assessment/ monitoring

To what extent is the data quality within [*your team(s)/departments*] effectively monitored with data quality tools (e.g., SAP Information Steward)?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective

If [CQ-1] = Data customer

[DQAM-2] Data quality assessment/ monitoring

The quality of the data supplied by [*the business chain*] is _____.

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Poor	Fair	Good	Very good	Excellent

Block 8. [CDQI] Continuous data quality (management) improvement

If [CQ-1] = Internal data supplier

[CDQI-1] Continuous improvement of data quality controls

(preventive, detective, corrective)

To what extent is [*your organizations*] focused on the continuous improvement of data quality controls?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Continuously improving **preventive** controls that prevent poor quality data from entering systems.	Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused
Continuously improving **detective** controls that detect poor quality data in the systems.	Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused
Continuously improving **corrective** controls that correct poor quality data in the systems.	Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused

[CDQI-2] Cost/benefit analyzes

To what extent are cost-benefit analyzes performed within [*your team(s)/departments*] before existing data quality controls are improved?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Never	Rarely	Sometimes	Often	Very often

[CDQI-3] Change management

To what extent is [*the organization*] able to effectively manage changes that influence data quality?

- Internal changes: e.g., organization restructuring, introducing new technology, personnel changes, culture changes, business process changes.
- External changes: e.g., government regulations, technology, economy, and market changes.

			Score = 1	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**
		Managing **internal** changes	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective
		Managing **external** changes	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective
Block 9. [DSQM] Data supplier quality management							
Only if the business chain has data suppliers.							
If [CQ-1] = External data supplier	[DSQM-1] Agreement about acceptable data quality level	To what extent do you agree with the following statement? There are clear agreements between [*the organization*] and [*your team(s)/ department(s)*] about the quality level that the supplied data must meet.					
		Score = 1	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**	
		Not clear at all	Slightly clear	Somewhat clear	Very clear	Extremely clear	
	[DSQM-2] Data quality reports and technical assistance	To what extent is [*your team(s)/ department(s)*] provided with (1) data quality reports and (2) technical assistance by [*your organization*]?					
			Score = 1	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**
		Data quality reports that gives [*your team(s)/department(s)*] insight into the quality level of the supplied data.	Never	Rarely	Sometimes	Often	Very often
		Technical assistance to improve data quality management within [*your team(s)/department(s)*].	Never	Rarely	Sometimes	Often	Very often
Block 10. [DSM] Data security management							
If [CQ-1] = Internal data supplier	[DSM-1] Access controls and security	To what extent do you agree with the following statement?					
			Score = 1	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**
		Users of information systems within [*your team(s)/department(s)*] are appropriately **authenticated** (e.g., using passwords, tokens).	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
		Users of information systems within [*your team(s)/department(s)*] are **authorized** with the least set of privileges they need to do their job.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
		To what extent are audit trails (logs of users' activities) of the information systems used within [*your team(s)/department(s)*] analyzed and periodically reviewed (e.g., for exceptions)?					
		Score = 1	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**	
		Never	Rarely	Sometimes	Often	Very often	
	[DSM-2] Human and process controls	To what extent do you agree with the following statement? Segregation of duties is ensured within [*the business chain*], i.e., having more than one person required to complete a critical task to prevent fraud and/or error.					
		Score = 1	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**	
		Strongly disagree	Disagree	Neutral	Agree	Strongly agree	
Block 11. [DPLM] Data product lifecycle management							
If [CQ-1] = Internal data supplier	[DPLM-1] Data product supply chain management	To what extent do you agree with the following statement? There are clear agreements within [*the business chain*] about the quality level that the data must meet.					
		Score = 1	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**	
		Strongly disagree	Disagree	Neutral	Agree	Strongly agree	
	[DPLM-2] Data migration	To what extent do you agree with the following statements? *ETL = Extraction, Transformation, and Load					

			Score = 1	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**
		Data migration (i.e., ETL) is minimized as much as possible within [*the business chain*].	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
		When migration data (i.e., ETL) within [*the business chain*], standard methods for system and software development are followed and extensive testing is included.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
	[DPLM-3] Metadata management	To what extent is metadata managed appropriately within [*your organization*]?					
		Score = 1	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**	
		Not appropriate at all	Slightly appropriate	Somewhat appropriate	Very appropriate	Extremely appropriate	
Block 12. [DAM] Data architecture management							
If [CQ-1] = Internal data supplier	[DAM-1] Data integration	To what extent has the integration of data (i.e., combining data from different sources into a single, unified view) within [*the business chain*] been maximized?					
		Score = 1	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**	
		Not maximized at all	Slightly maximized	Somewhat maximized	Very maximized	Extremely maximized	
	[DAM-2] Interface management	To what extent have the number of interfaces (i.e., links between different systems) within [*the business chain*] been minimized as much as possible?					
		Score = 1	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**	
		Not minimized at all	Slightly minimized	Somewhat minimized	Very minimized	Extremely minimized	

Appendix IV –Blueprint of the assessment instrument (version II)

Welcome!

Thank you for taking the time to participate in this survey to factors influencing data quality management within:

'[*the business chain*]'

[ADD EXPLANATION ON SCOPE OF BUSINESS CHAIN]

Participation in this survey allows you to give your opinion about your experiences with critical success factors that influence data quality management within the chain. The data you leave will be used to make improvement within the chain. You are completely anonymous while completing this survey.

Contingency questions

Logic	Code	Questions and answer options
	[CQ-1]	What is your role within [*the business chain*]? Internal data supplier | External data supplier | Data customer
If [CQ-1] = Internal data supplier	[TD-a]	Select one or more teams/departments you are part of below: *[Team/department name 1]* | *[Team/department name 4]* | *[Team/department name 7]* *[Team/department name 2]* | *[Team/department name 5]* | *[Team/department name 8]* *[Team/department name 3]* | *[Team/department name 6]* | *[Team/department name N]*
If [CQ-1] = External data supplier	[TD-b]	Select one or more teams/departments you are part of below: *[Team/department name 1]* | *[Team/department name 2]* | *[Team/department name N]*
If [CQ-1] = Data customer	[TD-c]	Select one or more teams/departments you are part of below: *[Team/department name 1]* | *[Team/department name 3]* | *[Team/department name 5]* *[Team/department name 2]* | *[Team/department name 4]* | *[Team/department name N]*

Questions that assesses the domain reference model

Block 1. [MCLS] Management commitment, leadership, and support

For the questions below, *management* is not confined to top or senior management, but refers to all levels of management within [*the organization*].

If [CQ-1] = Internal data supplier

[MCLS-1] Recognition of importance and responsibility

To what extent do you agree with the following statements?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Management [*within the organization*] recognizes the importance of data quality.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Management [*within the organization*] takes responsibility for the quality of the data.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

[MCLS-2] Allocation of resources

To what extent does management allocates sufficient resources to support data quality within [*your team(s)/department(s)*]?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Sufficient **budget**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Sufficient **technical tools (software)**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Sufficient **expertise (knowledge)**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Sufficient **skilled personnel (skills)**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Sufficient **time**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

[MCLS-3] Attribution of rewards

How often does management attributes rewards to employees within [*your team(s)/department(s)*] who are proactively committed to managing data quality?

- Examples of rewards: recognition for data quality improvement suggestions; increased budget; positive feedback.

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Never	Rarely	Sometimes	Often	Very often

[MCLS-4] Strategic data quality plan

To what extent do you agree with the following statement?
There is a clear plan for improving data quality within the systems managed by [*your team(s)/department(s)*].

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Strongly disagree	Disagree	Neutral	Agree	Strongly agree

Block 2. [DG] Data governance

If [CQ-1] = Internal data supplier

[DG-1] Organizational structure

To what extent do you agree with the following statements?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
The organizational structure within [*the organization*] is suitable to produce high-quality data.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
There are clear roles and responsibilities regarding data quality.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

[DG-2] Teamwork

To what extent is there effective teamwork between business and IT people?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Teamwork **within** [*your team(s)/department(s)*]	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective
Teamwork **between** [*your team(s)/department(s)*] and other teams/departments within [*the business chain*]	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective

Block 3. [CC] Culture and communication

If [CQ-1] = Internal data supplier

[CC-1] Organizational culture focused on data quality

To what extent does [*the organization*] have a culture that focuses on the continuous and proactive management of data quality?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused

[CC-2] Communication

To what extent is there effective communication between business and IT people?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Communication **within** [*your team(s)/department(s)*]	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective
Communication **between** [*your team(s)/department(s)*] and other teams/departments within [*the business chain*]	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective

<table>
<tr><td></td><td>[CC-3]
Knowledge sharing</td><td>How often is knowledge shared within [the organization] at knowledge sharing occasions and via networks of experts?
<table><tr><th></th><th>Score = 1</th><th>Score = 2</th><th>Score = 3</th><th>Score = 4</th><th>Score = 5</th></tr><tr><td>Regular knowledge sharing (repeated at specific intervals, e.g. once every month)</td><td>Never</td><td>Rarely</td><td>Sometimes</td><td>Often</td><td>Very often</td></tr><tr><td>Event-triggered knowledge sharing (at specific events, e.g., project’s ends, coming up of a new technology)</td><td>Never</td><td>Rarely</td><td>Sometimes</td><td>Often</td><td>Very often</td></tr><tr><td>Network of experts (e.g., knowledge teams or centers, communities of practice)</td><td>Never</td><td>Rarely</td><td>Sometimes</td><td>Often</td><td>Very often</td></tr></table></td></tr>
<tr><td colspan="3">Block 4. [ET] Education and training</td></tr>
<tr><td rowspan="2">If [CQ-1] = Internal data supplier</td><td>[ET-1]
Training workshops/ programs</td><td>To what extent does [the organization] offer sufficient training on data quality management?
<table><tr><th></th><th>Score = 1</th><th>Score = 2</th><th>Score = 3</th><th>Score = 4</th><th>Score = 5</th></tr><tr><td>Sufficient initial training on data quality management (new personnel, new/ upgrade system).</td><td>Strongly disagree</td><td>Disagree</td><td>Neutral</td><td>Agree</td><td>Strongly agree</td></tr><tr><td>Sufficient ongoing training (apart from the initial training) on data quality management offered (regular training for employees/managed, e.g. once every year).</td><td>Strongly disagree</td><td>Disagree</td><td>Neutral</td><td>Agree</td><td>Strongly agree</td></tr></table></td></tr>
<tr><td>[ET-2]
Professional development</td><td>To what extent is on-the-job professional development of employees supported by [the organization] (e.g., by providing mentoring programs)?
<table><tr><th>Score = 1</th><th>Score = 2</th><th>Score = 3</th><th>Score = 4</th><th>Score = 5</th></tr><tr><td>Not supported at all</td><td>Slightly supported</td><td>Somewhat supported</td><td>Very supported</td><td>Extremely supported</td></tr></table></td></tr>
<tr><td colspan="3">Block 5. [DQREM] Data quality requirements management</td></tr>
<tr><td rowspan="2">If [CQ-1] = Internal data supplier</td><td>[DQREM-1a]
Stakeholder management</td><td>To what extent is data quality central to prioritizing the requests that [your team(s)/department(s)] realizes?
<table><tr><th>Score = 1</th><th>Score = 2</th><th>Score = 3</th><th>Score = 4</th><th>Score = 5</th></tr><tr><td>Not central at all</td><td>Slightly central</td><td>Somewhat central</td><td>Very central</td><td>Extremely central</td></tr></table></td></tr>
<tr><td>[DQREM-2]
Data (quality) requirements management</td><td>To what extent do you agree with the following statements?
<table><tr><th></th><th>Score = 1</th><th>Score = 2</th><th>Score = 3</th><th>Score = 4</th><th>Score = 5</th></tr><tr><td>Data requirements are accurately captured within [your team(s) /departments] by modeling them in conceptual, logical, and physical data models.</td><td>Strongly disagree</td><td>Disagree</td><td>Neutral</td><td>Agree</td><td>Strongly agree</td></tr><tr><td>Data quality requirements are proactively identified, verified, validated, and updated within [your team(s)/departments] in consultation with stakeholders.</td><td>Strongly disagree</td><td>Disagree</td><td>Neutral</td><td>Agree</td><td>Strongly agree</td></tr></table></td></tr>
<tr><td>f [CQ-1] = Data customers</td><td>[DQREM-1b]
Stakeholder management</td><td>To what extent is [the business chain] focused on the needs and quality requirements that [your team(s)/ department(s)] places on the supplied data?
<table><tr><th></th><th>Score = 1</th><th>Score = 2</th><th>Score = 3</th><th>Score = 4</th><th>Score = 5</th></tr><tr><td>Needs of [your team(s)/ department(s)]</td><td>Not focused at all</td><td>Slightly focused</td><td>Somewhat focused</td><td>Very focused</td><td>Extremely focused</td></tr><tr><td>Quality requirements that [your team(s)/ department(s)] places on the supplied data.</td><td>Not focused at all</td><td>Slightly focused</td><td>Somewhat focused</td><td>Very focused</td><td>Extremely focused</td></tr></table></td></tr>
</table>

Block 6. [DQRIM] Data quality risk management

If [CQ-1] = Internal data supplier

[DQRIM-1] Awareness of consequences

To what extent is [*the organization*] aware of the consequences that poor data quality within [*the business chain*] can have?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not aware at all	Slightly aware	Somewhat aware	Very aware	Extremely aware

[DQRIM-2] Commitment to impact reducing

To what extent is [*the organization*] committed to reducing the impact that poor data quality within [*the business chain*] can have?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not committed at all	Slightly committed	Somewhat committed	Very committed	Extremely committed

Block 7. [DQAM] Data quality assessment/monitoring

If [CQ-1] = Internal data supplier

[DQAM-1] Data profiling

To what extent does [*your team(s)/departments*] make adequate use of data profiling to proactively identify data quality problems?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not adequate at all	Slightly adequate	Somewhat adequate	Very adequate	Extremely adequate

[DQAM-2] Data quality assessment/ monitoring

To what extent is the data quality within [*your team(s)/departments*] effectively monitored with data quality tools (e.g., SAP Information Steward)?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective

If [CQ-1] = Data customer

[DQAM-2] Data quality assessment/ monitoring

The quality of the data supplied by [*the business chain*] is _____.

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Poor	Fair	Good	Very good	Excellent

Block 8. [CDQI] Continuous data quality (management) improvement

If [CQ-1] = Internal data supplier

[CDQI-1] Continuous improvement of data quality controls

(preventive, detective, corrective)

To what extent is [*your team(s)/departments*] focused on the continuous improvement of data quality controls?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Continuously improving **preventive** controls that prevent poor quality data from entering systems managed by [*your team(s)/departments*].	Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused
Continuously improving **detective** controls that detect poor quality data in the systems managed by [*your team(s)/departments*].	Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused
Continuously improving **corrective** controls that correct poor quality data in the systems managed by [*your team(s)/departments*].	Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused

[CDQI-2] Cost/benefit analyzes

To what extent are cost-benefit analyzes performed within [*your team(s)/departments*] before existing data quality controls are improved?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Never	Rarely	Sometimes	Often	Very often

[CDQI-3]

To what extent is [*the organization*] able to effectively manage changes that influence data quality?

	Change management		**Score = 1**	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**
		Managing **internal** changes (e.g., organizational restructuring, changes in personnel / culture / business processes).	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective
		Managing **external** changes (e.g., government regulations, economy, technology, market changes).	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective
Block 9. [DSQM] Data supplier quality management							
Only if the business chain has data suppliers.							
If [CQ-1] = External data supplier	[DSQM-1] Agreement about acceptable data quality level	To what extent do you agree with the following statement? There are clear agreements between [*the organization*] and [*your team(s)/department(s)*] about the quality level that the supplied data must meet.	**Score = 1** Not clear at all	**Score = 2** Slightly clear	**Score = 3** Somewhat clear	**Score = 4** Very clear	**Score = 5** Extremely clear
	[DSQM-2] Data quality reports and technical assistance	To what extent is [*your team(s)/department(s)*] provided with (1) data quality reports and (2) technical assistance by [*your organization*]?	**Score = 1**	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**
		Data quality reports that gives [*your team(s)/department(s)*] insight into the quality level of the supplied data.	Never	Rarely	Sometimes	Often	Very often
		Technical assistance to improve data quality management within [*your team(s)/department(s)*].	Never	Rarely	Sometimes	Often	Very often
Block 10. [DSM] Data security management							
If [CQ-1] = Internal data supplier	[DSM-1] Access controls and security	To what extent do you agree with the following statement?	**Score = 1**	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**
		Users of information systems within [*your team(s)/department(s)*] are appropriately **authenticated** (e.g., using passwords, tokens).	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
		Users of information systems within [*your team(s)/department(s)*] are **authorized** with the least set of privileges they need to do their job.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
		To what extent are audit trails (logs of users' activities) of the information systems used within [*your team(s)/department(s)*] analyzed and periodically reviewed (e.g., for exceptions)?	**Score = 1** Never	**Score = 2** Rarely	**Score = 3** Sometimes	**Score = 4** Often	**Score = 5** Very often
	[DSM-2] Human and process controls	To what extent do you agree with the following statement? Segregation of duties is ensured within [*the business chain*], i.e., having more than one person required to complete a critical task to prevent fraud and/or error.	**Score = 1** Strongly disagree	**Score = 2** Disagree	**Score = 3** Neutral	**Score = 4** Agree	**Score = 5** Strongly agree
Block 11. [DPLM] Data product lifecycle management							
If [CQ-1] = Internal data	[DPLM-1] Data product supply chain management	To what extent do you agree with the following statement? There are clear agreements within [*the business chain*] about the quality level that the data must meet.	**Score = 1** Strongly disagree	**Score = 2** Disagree	**Score = 3** Neutral	**Score = 4** Agree	**Score = 5** Strongly agree

	[DPLM-2] Data migration	To what extent do you agree with the following statements? *ETL = Extraction, Transformation, and Load

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Data migration (i.e., ETL) is minimized as much as possible within *[the business chain*].	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
When migration data (i.e., ETL) within [*the business chain*], standard methods for system and software development are followed and extensive testing is included.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

[DPLM-3]
Metadata management

To what extent is metadata managed appropriately within [*your team(s)/department(s)*]?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not appropriate at all	Slightly appropriate	Somewhat appropriate	Very appropriate	Extremely appropriate

Block 12. [DAM] Data architecture management

If [CQ-1] = Internal data supplier

[DAM-1]
Data integration

To what extent has the integration of data (i.e., combining data from different sources into a single, unified view) within [*the business chain*] been maximized?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not maximized at all	Slightly maximized	Somewhat maximized	Very maximized	Extremely maximized

[DAM-2]
Interface management

To what extent have the number of interfaces (i.e., links between different systems) within [*the business chain*] been minimized as much as possible?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not minimized at all	Slightly minimized	Somewhat minimized	Very minimized	Extremely minimized

Appendix V – Blueprint of the assessment instrument (version III)

Welcome!

Thank you for taking the time to participate in this survey to factors influencing data quality management within:

'[*the business chain*]'

[ADD EXPLANATION ON SCOPE OF BUSINESS CHAIN]

Participation in this survey allows you to give your opinion about your experiences with critical success factors that influence data quality management within the chain. The data you leave will be used to make improvement within the chain. You are completely anonymous while completing this survey.

Contingency questions

Logic	Code	Questions and answer options
	[CQ-1]	What is your role within [*the business chain*]? Internal data supplier | External data supplier | Data customer
If [CQ-1] = Internal data supplier	[TD-a]	Select one or more teams/departments you are part of below: *[Team/department name 1]* | *[Team/department name 4]* | *[Team/department name 7]* *[Team/department name 2]* | *[Team/department name 5]* | *[Team/department name 8]* *[Team/department name 3]* | *[Team/department name 6]* | *[Team/department name N]*
If [CQ-1] = External data supplier	[TD-b]	Select one or more teams/departments you are part of below: *[Team/department name 1]* | *[Team/department name 2]* | *[Team/department name N]*
If [CQ-1] = Data customer	[TD-c]	Select one or more teams/departments you are part of below: *[Team/department name 1]* | *[Team/department name 3]* | *[Team/department name 5]* *[Team/department name 2]* | *[Team/department name 4]* | *[Team/department name N]*

Questions that assesses the domain reference model

Block 1. [MCLS] Management commitment, leadership, and support

For the questions below, *management* is not confined to top or senior management, but refers to all levels of management within [*the organization*].

If [CQ-1] = Internal data supplier

[MCLS-1] Recognition of importance and responsibility

To what extent do you agree with the following statements?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Management [*within the organization*] recognizes the importance of data quality.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Management [*within the organization*] takes responsibility for the quality of the data.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

	[MCLS-2] Allocation of resources	To what extent does management allocates sufficient resources to support data quality within [*your team(s)/department(s)*]?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Sufficient **budget**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Sufficient **technical tools (software)**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Sufficient **expertise (knowledge)**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Sufficient **skilled personnel (skills)**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Sufficient **time**	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

[MCLS-3] Attribution of rewards

How often does management attributes rewards to employees within [*your team(s)/department(s)*] who are proactively committed to managing data quality?

- Examples of rewards: recognition for data quality improvement suggestions; increased budget; positive feedback.

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Never	Rarely	Sometimes	Often	Very often

[MCLS-4] Strategic data quality plan

To what extent do you agree with the following statement?
There is a clear plan for improving data quality within the systems managed by [*your team(s)/department(s)*].

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Strongly disagree	Disagree	Neutral	Agree	Strongly agree

Block 2. [DG] Data governance

If [CQ-1] = Internal data supplier

[DG-1] Organizational structure

To what extent do you agree with the following statements?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
The organizational structure within [*the organization*] is suitable to produce high-quality data.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
There are clear roles and responsibilities regarding data quality.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

[DG-2] Teamwork

To what extent is there effective teamwork between business and IT people?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Teamwork **within** [*your team(s)/department(s)*]	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective
Teamwork **between** [*your team(s)/department(s)*] and other teams/departments within [*the business chain*]	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective

Block 3. [CC] Culture and communication

If [CQ-1] = Internal data supplier

[CC-1] Organizational culture focused on data quality

To what extent does [*the organization*] have a culture that focuses on the continuous and proactive management of data quality?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused

[CC-2] Communication

To what extent is there effective communication between business and IT people?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Communication **within** [*your team(s)/department(s)*]	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective
Communication **between** [*your team(s)/department(s)*] and other teams/departments within [*the business chain*]	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective

	[CC-3] Knowledge sharing	How often is knowledge shared within [*the organization*] at knowledge sharing occasions and via networks of experts? (not limited to data quality)

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Regular knowledge sharing (repeated at specific intervals, e.g. once every month)	Never	Rarely	Sometimes	Often	Very often
Event-triggered knowledge sharing (at specific events, e.g., project's ends, coming up of a new technology)	Never	Rarely	Sometimes	Often	Very often
Network of experts (e.g., knowledge teams or centers, communities of practice)	Never	Rarely	Sometimes	Often	Very often

Block 4. [ET] Education and training

If [CQ-1] = Internal data supplier

[ET-1] Training workshops/ programs

To what extent does [*the organization*] offer sufficient training on data quality management?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Sufficient **initial training** on data quality management (new personnel, new/ upgrade system).	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Sufficient **ongoing training (apart from the initial training)** on data quality management offered (regular training for employees/managed, e.g. once every year).	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

[ET-2] Professional development

To what extent is on-the-job professional development of employees supported by [*the organization*] (e.g., by providing mentoring programs)?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not supported at all	Slightly supported	Somewhat supported	Very supported	Extremely supported

Block 5. [DQREM] Data quality requirements management

If [CQ-1] = Internal data supplier

[DQREM-1a] Stakeholder management

To what extent is data quality central to prioritizing the requests that [*your team(s)/department(s)*] realizes?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not central at all	Slightly central	Somewhat central	Very central	Extremely central

[DQREM-2] Data (quality) requirements management

To what extent do you agree with the following statements?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Data requirements are accurately captured within [*your team(s) /departments*] by modeling them in conceptual, logical, and physical data models.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Data quality requirements are proactively identified, verified, validated, and updated within [*your team(s)/departments*] in consultation with stakeholders.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

If [CQ-1] = Data customers

[DQREM-1b] Stakeholder management

To what extent is [*the business chain*] focused on the needs and quality requirements that [*your team(s)/ department(s)*] places on the supplied data?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Needs of [*your team(s)/ department(s)*]	Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused
Quality requirements that [*your team(s)/ department(s)*] places on the supplied data.	Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused

Block 6. [DQRIM] Data quality risk management

If [CQ-1] = Internal data supplier

[DQRIM-1] Awareness of consequences

To what extent is [*the organization*] aware of the consequences that poor data quality within [*the business chain*] can have?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not aware at all	Slightly aware	Somewhat aware	Very aware	Extremely aware

[DQRIM-2] Commitment to impact reducing

To what extent is [*the organization*] committed to reducing the impact that poor data quality within [*the business chain*] can have?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not committed at all	Slightly committed	Somewhat committed	Very committed	Extremely committed

Block 7. [DQAM] Data quality assessment/monitoring

If [CQ-1] = Internal data supplier

[DQAM-1] Data profiling

To what extent does [*your team(s)/departments*] make adequate use of data profiling to proactively identify data quality problems?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not adequate at all	Slightly adequate	Somewhat adequate	Very adequate	Extremely adequate

[DQAM-2] Data quality assessment/ monitoring

To what extent is the data quality within [*your team(s)/departments*] effectively monitored with data quality tools (e.g., SAP Information Steward)?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective

If [CQ-1] = Data customer

[DQAM-2] Data quality assessment/ monitoring

The quality of the data supplied by [*the business chain*] is _____.

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Poor	Fair	Good	Very good	Excellent

Block 8. [CDQI] Continuous data quality (management) improvement

If [CQ-1] = Internal data supplier

[CDQI-1] Continuous improvement of data quality controls

(preventive, detective, corrective)

To what extent is [*your team(s)/departments*] focused on the continuous improvement of data quality controls?

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Continuously improving **preventive** controls that prevent poor quality data from entering systems managed by [*your team(s)/departments*].	Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused
Continuously improving **detective** controls that detect poor quality data in the systems managed by [*your team(s)/departments*].	Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused
Continuously improving **corrective** controls that correct poor quality data in the systems managed by [*your team(s)/departments*].	Not focused at all	Slightly focused	Somewhat focused	Very focused	Extremely focused

[CDQI-2] Cost/benefit analyzes

To what extent are cost-benefit analyzes performed within [*your team(s)/departments*] before existing data quality controls are improved?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Never	Rarely	Sometimes	Often	Very often

[CDQI-3]

To what extent is [*the organization*] able to effectively manage changes that influence data quality?

	Change management		**Score = 1**	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**
		Managing **internal** changes (e.g., organizational restructuring, changes in personnel / culture / business processes).	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective
		Managing **external** changes (e.g., government regulations, economy, technology, market changes).	Not effective at all	Slightly effective	Somewhat effective	Very effective	Extremely effective
Block 9. [DSQM] Data supplier quality management							
Only if the business chain has data suppliers.							
If [CQ-1] = External data supplier	[DSQM-1] Agreement about acceptable data quality level	To what extent do you agree with the following statement? There are clear agreements between [*the organization*] and [*your team(s)/department(s)*] about the quality level that the supplied data must meet.	**Score = 1** Not clear at all	**Score = 2** Slightly clear	**Score = 3** Somewhat clear	**Score = 4** Very clear	**Score = 5** Extremely clear
	[DSQM-2] Data quality reports and technical assistance	To what extent is [*your team(s)/department(s)*] provided with (1) data quality reports and (2) technical assistance by [*your organization*]?	**Score = 1**	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**
		Data quality reports that gives [*your team(s)/department(s)*] insight into the quality level of the supplied data.	Never	Rarely	Sometimes	Often	Very often
		Technical assistance to improve data quality management within [*your team(s)/department(s)*].	Never	Rarely	Sometimes	Often	Very often
Block 10. [DSM] Data security management							
If [CQ-1] = Internal data supplier	[DSM-1] Access controls and security	To what extent do you agree with the following statement?	**Score = 1**	**Score = 2**	**Score = 3**	**Score = 4**	**Score = 5**
		Users of information systems within [*your team(s)/department(s)*] are appropriately **authenticated** (e.g., using passwords, tokens).	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
		Users of information systems within [*your team(s)/department(s)*] are **authorized** with the least set of privileges they need to do their job.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
		To what extent are audit trails (logs of users' activities) of the information systems used within [*your team(s)/department(s)*] analyzed and periodically reviewed (e.g., for exceptions)?	**Score = 1** Never	**Score = 2** Rarely	**Score = 3** Sometimes	**Score = 4** Often	**Score = 5** Very often
	[DSM-2] Human and process controls	To what extent do you agree with the following statement? Having more than one person required to complete a critical task to prevent fraud and/or error is ensured within [*your team(s)/department(s)*].	**Score = 1** Strongly disagree	**Score = 2** Disagree	**Score = 3** Neutral	**Score = 4** Agree	**Score = 5** Strongly agree
Block 11. [DPLM] Data product lifecycle management							
If [CQ-1] = Internal data	[DPLM-1] Data product supply chain management	To what extent do you agree with the following statement? There are clear agreements within [*the business chain*] about the quality level that the data must meet.	**Score = 1** Strongly disagree	**Score = 2** Disagree	**Score = 3** Neutral	**Score = 4** Agree	**Score = 5** Strongly agree

	[DPLM-2] Data migration	To what extent do you agree with the following statements? *ETL = Extraction, Transformation, and Load

	Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Data migration (i.e., ETL) is minimized as much as possible within *[the business chain*].	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
When migration data (i.e., ETL) within [*the business chain*], standard methods for system and software development are followed and extensive testing is included.	Strongly disagree	Disagree	Neutral	Agree	Strongly agree

[DPLM-3]
Metadata management

To what extent is metadata managed appropriately within [*your team(s)/department(s)*]?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not appropriate at all	Slightly appropriate	Somewhat appropriate	Very appropriate	Extremely appropriate

Block 12. [DAM] Data architecture management

If [CQ-1] = Internal data supplier

[DAM-1]
Data integration

To what extent has the integration of data (i.e., combining data from different sources into a single, unified view) within [*the business chain*] been maximized?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not maximized at all	Slightly maximized	Somewhat maximized	Very maximized	Extremely maximized

[DAM-2]
Interface management

To what extent have the number of interfaces (i.e., links between different systems) within [*the business chain*] been minimized as much as possible?

Score = 1	Score = 2	Score = 3	Score = 4	Score = 5
Not minimized at all	Slightly minimized	Somewhat minimized	Very minimized	Extremely minimized