UTRECHT UNIVERSITY

MASTER'S THESIS

A Framework for Data Quality Management in the Delivery & Consultancy of CRM Platforms

Author:
Renee ALBRECHT

Supervisor:
Dr. Sietse J. OVERBEEK

Second Supervisor:
Dr. Inge VAN DE WEERD

External Supervisors:
Aschwin STACIA
Shanice BULK

A thesis submitted in fulfillment of the requirements for the degree of Master of Science in Business Informatics

August 1, 2021

UTRECHT UNIVERSITY

Abstract

Faculty of Science
Department of Information and Computing Sciences

Master of Science

A Framework for Data Quality Management in the Delivery & Consultancy of CRM Platforms

by Renee ALBRECHT

Contemporary enterprises depend heavily on data. An important operational aspect of enterprises that relies upon its utilisation of high-quality data is CRM, where poor quality of data can negatively influence its adoption. Furthermore, contemporary CRM platforms are increasingly interconnected and complex due to varying and growing needs of customers. Hence, delivery & consultancy of such CRM platforms becomes increasingly complex as well. This research proposes a CRM data quality management framework that can assist CRM delivery & consultancy teams to improve data quality management practices within their projects and ultimately improve data quality within CRM solutions for their clients. The main research question reads as follows: How can a data quality management framework be designed to assist with the delivery and consultancy of CRM platforms? To answer the main research question, a literature study investigates data quality definition and measurement in the contect of CRM, existing data quality challenges in CRM platforms and their potential solutions, and existing data quality management practices, from which best practices for the context of CRM delivery & consultancy are extracted. A case study comprising expert interviews and documentation analysis at an IT consulting company investigates CRM delivery & consultancy projects and how these can benefit from the incorporation of data quality management practices. The design of the framework is validated by means of a design theory and a questionnaire, which are discussed in confirmatory mini focus groups consisting of CRM delivery & consultancy experts. The results translate into a framework that provides a high-level overview of data quality management practices incorporated in CRM delivery & consultancy projects. It involves the recognition of variety in clients and projects by introducing the establishment of a unique data quality management plan at the start of the project. Furthermore, it includes the following components: client profiling; project definition; preparation; migration/integration; data quality definition; assessment; and improvement.

Acknowledgements

First of all, I would like to thank my first supervisor, dr. Sietse Overbeek, for his advice and critical feedback about my work. His availability and fast responses are highly appreciated, and have truely helped me with the completion of this research. Therefore, a great thank you.

I would like to thank my supervisors from Avanade, Aschwin Stacia and Shanice Bulk, for their support during this research project. They arranged many interesting and instructive activities for me to take part in, guided me through the entire thesis process at Avanade, and provided critical feedback. Also, they provided for fun social talks on a weekly basis during this period of working remote due to the pandemic.

Furthermore, I want to thank dr. Inge van de Weerd for accepting to be my second supervisor and evaluating my work.

I would like to thank all participants of this research for their contribution. Their efforts to take part have been highly important to this research.

Lastly, I would like to thank Daan and Takis - my fellow students at Utrecht University - for their support during our thesis projects, and Rebecca and Tineke - my fellow Avanade interns - for the many updates and fun meetings.

Contents

List of Figures

List of Tables

List of Abbreviations

CRM	Customer Relationship Management.
CRM-DQMF	CRM DQ Management Framework.
D365	Microsoft Dynamics 365.
D&C	Delivery and Consultancy.
DM	Data Mining.
DQ	Data Quality.
ERP	Enterprise Resource Planning.
ETL	Extract, Transform, Load.
GDPR	General Data Protection Regulation.
IQ	Information Quality.
KPI	Key Performance Indicator.
PDD	Process-Deliverable Diagram.
RFM	Recency, Frequency, Monetary.
SaaS	Software as a Service.

Chapter 1

Introduction

1.1 Problem Statement

Contemporary enterprises depend heavily on data, where data is seen as a strategic asset (Nagle, Redman, & Sammon, 2020). International Data Corporation (IDC) states in the Global DataSphere that during 2020 the amount of data that is created, captured, and replicated across the world in 2020 is more than 59 zettabytes, and over the next five years the world will create more than three times the amount of data of the previous five years (IDC, 2020). This results in an increase in the number of errors in data, such as information that is misplaced, fragmented or inconsistent. Poor Data Quality (DQ) is one of the greatest challenges facing contemporary enterprises (Davenport & Harris, 2017). Simultaneously, enterprises struggle to address their data issues, while high-quality data is rather the exception than the rule (Nagle, Redman, & Sammon, 2020).

A DQ survey conducted by The Data Warehouse Institute (TDWI) that drew responses from 647 individuals in a range of positions, industries, and countries showed that almost half of all participating companies have no plan for managing DQ (Eckerson, 2002). IBM estimates the costs of poor DQ within the US economy alone to be $3.1 trillion per year (IBM, 2018), and according to Gartner, poor DQ results in the undermining of digital initiatives, weakened competitive standing, and an increase in customer distrust (Gartner, 2018). With the orders of magnitude increases in volume, velocity and variety of data, the issues arising from errors in data and the organizational impact of these issues are greatly amplified (Laney, 2018). Laney states that "poor data quality can have grave consequences, from strategic decisions that can lead to the death of a business to operation decisions that can lead to the death of individuals." and "[...] 40 percent of all failed business initiatives are a result of poor data quality" (Laney, 2018, pp. 246–247). By managing DQ, unwanted consequences can be prevented, and valuable insights can be discovered in regards to interactions with customers.

An important operational aspect of enterprises that relies upon its utilisation of high-quality data is Customer Relationship Management (CRM). Cruz-Jesus et al. (2019) found that poor DQ and integration can negatively influence the adoption of CRM. Additionally, survey data collected from about 300 organisations for the *State of CRM Data Management* report shows that 44% of its respondents estimate a loss in revenue as a result of poor quality CRM data, which ranges from 5%-20% of total revenue (Hanson, 2020). CRM is defined as "the core business strategy that aims to create and maintain profitable relationships with customers, by designing and delivering superior value proposition" (Buttle & Maklan, 2019, p. 21). It is enabled by information technology in the form of CRM platforms, at present often provided by IT consultancy firms as CRM (cloud) solutions. Those firms offer Delivery and Consultancy (D&C) of the CRM platform they provide to a variety of customers

from various industries. Those customers vary and grow in their needs, business processes, and goals. Consequently, the CRM platforms are increasingly interconnected and complex, resulting in a continuous need for more study in the area of the management of CRM platform development, implementation and marketing (Stone et al., 2017). This research' objective is to provide a solution for the need for data of adequate quality in increasingly complex CRM platforms. Its aim following the template of Wieringa (2014) reads as follows:

> *This research aims to improve* data quality within CRM platforms *by* designing a CRM DQ management framework *that* assists CRM D&C teams with the delivery and consultancy processes of CRM platforms *in order to* keep the quality of data on a level required for the specific organization.

1.2 Research Questions

To achieve its aim, the following main research question (MRQ) is formulated:

[MRQ] *How can a data quality management framework be designed to assist with the delivery and consultancy of CRM platforms?*

The answer to the [MRQ] is in the form of the CRM DQ Management Framework (CRM-DQMF). This framework functions as a guide for DQ management within a CRM platform D&C project. To address the [MRQ], the following sub-questions (RQs) in the context of CRM platforms are formulated. All RQs are answered by thorough literature research and case study methods.

[RQ1] *What does the delivery and consultancy of a contemporary CRM platform consist of?*
The goal of [RQ1] is to provide an overview of what the delivery and consultancy of typical contemporary CRM platforms consist of. This includes an understanding of internal data processes, and possible integrations and interconnections with related applications or systems. The overview presents the context for answering the following RQs.

[RQ2] *How can quality of CRM data be defined?*
The goal of [RQ2] is to define quality for CRM data, for which the context is drawn from the answer to [RQ1]. This is done by defining what types of data are concerned with CRM platforms. So-called DQ dimensions describe the requirements that concerned data need to fulfill to be of high-quality (Wang, 1998). The answer to this question serves as input for [RQ3].

[RQ3] *How can CRM data quality be assessed?*
Research on the level of DQ in a wide range of enterprises reveals that there is a lack of awareness of the actual DQ level, which enterprises need for the understanding of required improvements and government concerning DQ (Nagle, Redman, & Sammon, 2020). The goal of [RQ3] is to determine what DQ measurements are most applicable for the quality of data in the context of CRM platforms as defined by [RQ2].

[RQ4] *What are existing CRM data quality challenges and their potential solutions?*
The goal of [RQ4] is to investigate existing challenges concerning DQ in CRM, as well as their potential solutions. This provides insights in criteria for DQ management practices in CRM platforms.

[RQ5] *What are existing data quality management approaches suitable for CRM platforms?*
The goal of [RQ5] is to examine existing DQ management approaches and extract best practices for the context of CRM platforms.

1.3 Research Environment

This research presents a Master's Thesis project of the MSc Business Informatics at Utrecht University. The research is carried out in collaboration with Avanade Netherlands, which is founded as a joint venture between Microsoft and Accenture. Avanade is an IT consultancy firm offering consultancy and delivery of the Microsoft ecosystem. Avanade's vision is to "be the leading digital innovator, creating great experiences as we realize results for our clients and their customers through the power of people and the Microsoft ecosystem"[1]. Amongst other Microsoft solutions, Avanade offers the solution Microsoft Dynamics 365 (D365), a CRM and Enterprise Resource Planning (ERP) applications platform. D365 is offered on location, in the cloud, or as a hybrid system to various industries, such as retail, financial services, manufacturing and healthcare. The three main industries utilising Avanade's services are the *utility industry*, *finance industry*, and *product industry*. At Avanade, a case study is performed and the designed artifact is validated by practitioners.

1.4 Research Approach

This section describes the research methods that are used in this project. The research concerns a design problem, with the aim of investigation and design of an artifact in context, namely a framework for DQ management in CRM platforms. This type of research is called Design Science (Wieringa, 2014). Hence, to achieve its aim, the research is framed around the Design Science Framework as proposed by Wieringa. The iterative problem-solving methodology is adopted, where the aim is to solve a problem by creating a new artifact through iterative activities of investigation and design. Those activities are broken down into a set of tasks according to the structure of the so-called design cycle. The tasks include *Problem Investigation*, *Treatment Design*, and *Treatment Validation*. In Figure 1.1 the tasks can be seen, along with the methods and analysis techniques that are used during the research. The task *Treatment Implementation* that is displayed in the figure is not included within this research, as this is only included within the so-called engineering cycle, of which the design cycle is a subset. The engineering cycle is often carried out in long-term projects, while design science projects are restricted to the design cycle, since implementation in real life is not part of a research project (Wieringa, 2014). The different steps that are carried out within this project are elaborated on in the paragraphs below.

[1] https://www.avanade.com/en/about-avanade

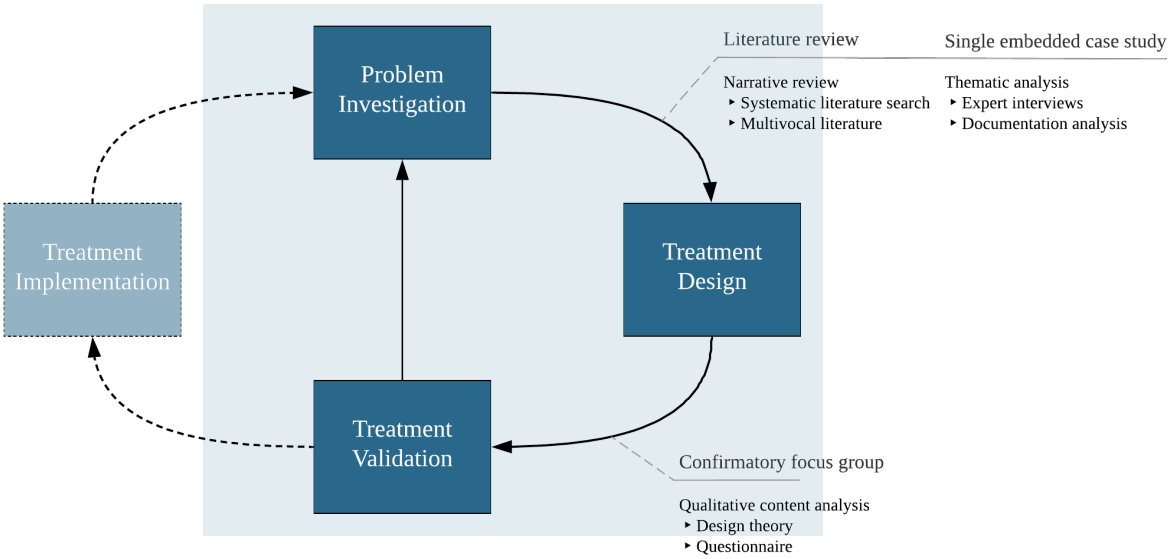

FIGURE 1.1: Design Cycle

Problem investigation: By learning more about the problem to be treated the design of a treatment is prepared. Stakeholders, desired goals, problems, phenomena, and effects are investigated to determine what should be improved and why by answering the RQs. Exploratory research is conducted to get an understanding of the problem. To come to this understanding related research from the past is examined (Webster & Watson, 2002) by conducting a systematic literature search (Kitchenham, 2004) as well as a multivocal literature review (Garousi, Felderer, & Mäntylä, 2016). Moreover, the literature review is extended to identify aspects of the problem by performing a single embedded case study, including semi-structured interviews with various stakeholders as well as documentation analysis.

Treatment design: Artifacts are designed to treat the problem. Hereby requirements are specified and the question whether they contribute to the research goal is answered. This is done by conducting literature research with the aim of defining high-quality data for CRM platforms ([RQ2]), and reviewing existing DQ management methods and their application to CRM ([RQ5]). Moreover, the case study is conducted to discover additional criteria for DQ and its management in CRM platforms (Runeson & Höst, 2007).

Treatment Validation: The effects of the designed artifact on the context are investigated and the question whether requirements are met is answered. The designed framework is validated by the opinions of experts within the case study (Wieringa, 2014).

In order to increase the robustness of the results of this research, methodological triangulation is adopted (Kaplan & Maxwell, 2005). The research triangulates by applying various methods for data collection, namely a systematic literature review including multivocal literature, and a single embedded case study including expert

interviews, documentation analysis, and expert opinions. The research methods are elaborated on in the following.

1.4.1 Systematic Literature Review

Data and CRM platforms are being utilised for numerous years; thus an extensive amount of studies have been performed on the topic of data quality methods and CRM practices. By performing a systematic literature review, findings of all relevant, high-quality studies that address one or more RQs are identified, critically evaluated, and integrated within this study (Baumeister & Leary, 1997). The systematic literature review is performed by conducting a systematic literature search, followed by a multivocal literature review.

Systematic Literature Search Protocol

The systematic literature search is carried out using a keyword-search in various scholarly databases, namely Google Scholar, ResearchGate, and the Computer Science Bibliography (DBLP). Keywords that are used are: *CRM, CRM platform, CRM AND data, CRM AND data quality, data quality management, data quality AND metrics, data quality AND dimensions, data quality method*. Selected document types are books, journals, conference proceedings, and theses. All literature is required to be in the language English and available in digital format. The relevancy of the literature is subjectively determined by scope, objectives, methods, and conclusion (Budgen & Brereton, 2006). To ensure a complete review of all relevant literature, additional search methods that are used are forward as well as backward searching (Levy & Ellis, 2006). By performing forward searching, literature that cites the already reviewed articles is reviewed as well, which results in the identification of more recent work. By performing backward searching, research on which found articles are based is reviewed.

Multivocal Literature Review

Contemporary scientific research is limited to the separate topics of DQ and CRM platforms, and there exists no extensive scientific research with the latest knowledge on the application of DQ management methods within contemporary CRM platforms (Wijnhoven et al., 2007). Therefore, a multivocal literature review approach is adopted to be able to review grey literature, meaning the literature is not peer-reviewed, and get insights on the latest state of DQ management in CRM platforms (Garousi, Felderer, & Mäntylä, 2016).

1.4.2 Case Study

For this research, a single embedded case study is adopted in regards to how an IT consultancy firm considers DQ in a CRM platform and assists other organisations with the adoption of said CRM platform. It allows for the capturing of the DQ management and CRM phenomena in general as well as the adoption of the artifact in practice (Benbasat, Goldstein, & Mead, 1987). Case study can be defined as an empirical method aimed at investigating contemporary phenomena in their context (Runeson & Höst, 2007; Yin, 2013; Robson & McCartan, 2016). It "explores a real-life, contemporary bounded system (a case) or multiple bounded systems (cases) over time, through detailed, indepth data collection involving multiple sources of information… and reports a case description and case themes." (Creswell & Poth,

2018, p. 96). In order to facilitate triangulation, several case study methods are applied within this study, namely systematising expert interviews and documentation analysis. Within the case study the opinions of participants are required for the design of the artifact. Thus, it can be considered interpretive, as it attempts to understand phenomena through the participants' interpretation of their context (Klein & Myers, 1999). The unit of analysis is CRM platforms, of which a sample containing one CRM platform is taken to perform the case study on. For the selection of participants of the case study, a non-probability sampling technique is used (Yin, 2013). The selection depends on various factors. The first factor is purposive, where the participants are selected because their views are relevant to the topic, which includes participants with specialist knowledge, participants in certain positions in the organisation, and the technique of snowballing, where new participants might be suggested by existing participants. The second factor is quota, since there are several sub-groups within the population of the case study, each with varying expertise of the different industries to which an IT consultancy firm might provide service. Every sub-group needs to be represented adequately.

Expert Interviews

Within the case study, interviews are conducted with several stakeholders as a data collection vehicle. The interviews are exploratory in nature, since their aim is to understand the field and identify design opportunities. A semi-structured approach is applied, because it facilitates probing questions depending on the specific interview (Kvale, 1996). Questions are prepared beforehand, but rather function as a guidance. In order to prevent maturation, the interviews are limited to 45 minutes up to one hour. The design of the interviews adheres to nine types of questions designed by Kvale (1996). This includes the following: introductory questions to gain background knowledge; structuring questions to structure the interview; follow-up questions based on a previous answer; probing questions to gather in depth information; specifying questions to request factual information and stay in scope; direct questions to directly introduce topics of interest; indirect questions looking for e.g. opinions or experiences; silence to let the interviewee think and possibly follow up with additional information; and interpreting questions to inspect whether the interviewee's answer is interpreted correctly. The interview protocol can be found in Appendix B.

Documentation analysis

Documentation analysis is considered a suitable data collection method within case study research. Documents are reviewed according to a systematic procedure (Bowen, 2009). It establishes facts, provides background to the case and inferential information on the studied CRM platform and its data. The documentation reports on the CRM platform, data processes, and existing DQ understanding related to the CRM platform. Documents can be, for example, manuals, background papers, guidelines, books, or agendas (Bowen, 2009). For this research, mainly company documentation will be utilised, which is private documentation. Since this means that the documentation is retrospectively analysed, the credibility and representativeness of the documents should be evaluated (Bryman, 2016). This is reported on by disclosing available metadata on the purpose of the document, the author type, document type, and the year of creation.

1.4.3 Design

The results that emerge from the aforementioned analysis and methods are presented by means of a Process-Deliverable Diagram (PDD) (Weerd & Brinkkemper, 2008; Weerd, 2009). A PDD is an assembly-based method engineering approach supported by a meta-modeling technique for the construction of situational analysis and design methods. It is based on UML activity and class diagrams. Method engineering is defined as "the engineering discipline to design, construct and adapt methods, techniques and tools for the development of information systems" (Brinkkemper, 1996). Since situational factors play a key role when managing DQ, (e.g. the CRM platform, the industry, the data processes etc.) situational method engineering is a suitable approach for the design of a DQ management CRM-DQMF. A PDD consists of two integrated diagrams, namely a process diagram including all activities, and a deliverable diagram including the deliverables that result from the activities.

1.4.4 Validation Strategy

The designed CRM-DQMF and underlying perspectives need to be validated to ensure credibility of the artifacts. By utilising the expert opinion approach the artifacts are validated. The design is applied to a simulated problem context to establish a design theory (Wieringa, 2014). Subsequently, an interactive questionnaire is conducted with experts, as this is perceived as an effective and one of the most common approaches (Robson & McCartan, 2016). A more elaborate explanation on the validation strategy can be foung in Chapter 5.

Expert Opinion

By applying the expert opinion approach, the artifact is subjected to and validated by experts. Those experts "imagine how such an artifact will interact with problem contexts imagined by them and then predict what effects they think this would have. If the predicted effects do not satisfy requirements, this is a reason to redesign the artifact." (Wieringa, 2014, p. 63). Opinions of experts are extracted by means of confirmatory focus groups (Tremblay, Hevner, & Berndt, 2010).

1.4.5 Data Analysis

The collected data is analysed in order to provide answers to the RQs. For the (multivocal) literature review, a narrative review method with the aim of reviewing the relevant sources and critically comparing existing DQ management methods and their possible application to CRM platforms (Baumeister, 2013).

For the analysis of the case study, thematic analysis method is applied (Clarke & Braun, 2014). A thematic analysis strives to identify patterns in the case study data. It is a qualitative method to extract themes from data with an active role for the researcher. A more elaborate explanation of the execution of thematic analysis can be found in Appendix B.

1.4.6 Research Summary

The individual actions and deliverables within the research process are visualized in a PDD that can be seen in Figure 1.2 (Weerd & Brinkkemper, 2008). The process is visualized within the respecting phases.

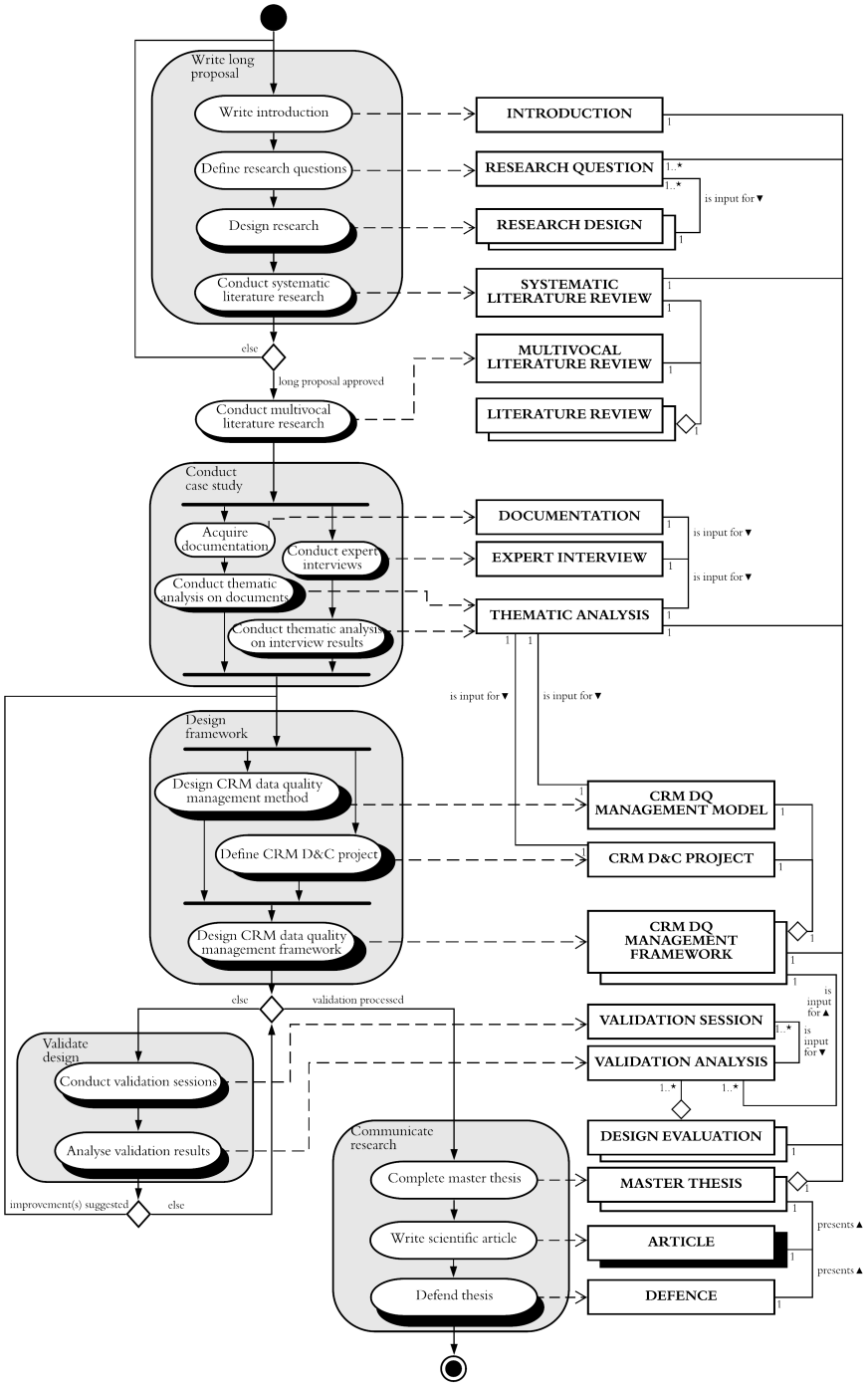

FIGURE 1.2: Research Overview

Chapter 2

Literature Review

2.1 Customer Relationship Management

This section provides an answer derived from literature to [RQ1]. CRM platforms provided by IT consultancy firms are elaborated on, as well as their D&C practices.

2.1.1 CRM Platforms

CRM is a strategic approach to systematically target, monitor, communicate and transform relevant customer data into information that underlies strategic decision making and action (Missi, Alshawi, & Fitzgerald, 2005). A CRM platform is a CRM software package that can be utilised by organisations to systematise their CRM (Cricelli et al., 2019). Software as a Service (SaaS) firms also offer CRM platforms in the cloud, meaning the delivery of computing services is over the internet instead of on-premise alternatives. The SaaS firm manages all aspects of the CRM platform's application environment, including the underlying infrastructure. It provides delivery of the CRM platform as well as consultancy, offering its customers CRM-related capabilities such as strategy, business, application and technical consulting for the most beneficial utilisation of the CRM platform (Buttle & Maklan, 2019). Customers wishing to implement a CRM platform in the cloud are able to utilise its functionalities by uploading their data to the host's servers and interact with the data using a web browser (Buttle & Maklan, 2019).

For CRM to be successful, business processes, human factor, and technology need to be integrated (Negahban, Kim, & Kim, 2016). The utilisation of a CRM platform depends on the precise requirements and specifications of an organisation. To meet those needs, contemporary CRM platforms offer an increasing amount of characteristics and possibilities to a variety of industries and are desired to be tailored for the specific organisation, require limited economic and human resources' effort, and be promptly deployed by specifically trained employees (Cricelli et al., 2019). CRM platforms can store information about customers and tailor relationships between the organisation and its customers (Nguyen & Mutum, 2012), and offer functionalities in the area of sales management, marketing, and customer service (Buttle & Maklan, 2019). Hence, the main characteristics of a CRM platform consist of a key focalisation of customer relationships by means of functional properties; power to integrate with and manage technology-based applications; and a wide capability of managing internal organisation activities and knowledge exchange by means of quality properties (Cricelli et al., 2019).

2.1.2 CRM Platform Delivery & Consultancy

When implementing a CRM platform, people, processes, and technology are crucial components (Ranki, 2019). People need to be convinced about the change as well as educated about the new CRM platform, as the successfulness of a CRM platform implementation depends on the skill of the human resources using it (Cruz-Jesus, Pinheiro, & Tiago, 2019). Processes need cross-process collaboration and restructuring, technology needs to be approachable, and available data sources need cleansing and integrating. The implementation of CRM platforms is often composed of a large number of smaller projects. Examples of projects can involve DQ improvement, market segmentation, process engineering and culture change (Buttle & Maklan, 2019).

For the delivery of a new CRM platform, the data that will be utilised need to be migrated from the original CRM system's data sources to the new platform data sources (Ali, Moawad, & Hosni, 2017). Data migration is defined as the process of permanently moving data from legacy data sources of a legacy system into new data sources of a target system, in which legacy and new systems often have different data structures (Thalheim & Wang, 2013). The legacy and new systems can either be an in-house on-premise system or a cloud solution. Hence, different migration scenarios can be required (Hai & Sakoda, 2009): on-premise to cloud; cloud to cloud; cloud to on-premise; or on-premise to on-premise. CRM platform providers often offer migration as part of their services by providing an integration server with prebuilt connectors to common on-premise applications or cloud-based CRM solutions (Hai & Sakoda, 2009; Ahmad, Naveed, & Hoda, 2018). However, those standard processes may not completely fit into the needs of the specific organisation (Birthare & Sharma, 2020). Therefore, custom solutions for the migration of data might be required. To do so, a mapping model should be developed by defining a profile of the new CRM platform as well as the legacy CRM platform, as the entity names of data can differ between the legacy platform and the new CRM platform. Data migration conventionally consists of multiple steps following the Extract, Transform, Load (ETL) processes (Thalheim & Wang, 2013). In *Extract* the data is extracted from the legacy data sources. In *Transform*, the data is transformed to meet the requirements of the new CRM platform. Migration transformations that are specified may involve validating, cleansing, and mapping the data. In *Load*, the data is loaded into the targeted data sources of the new platform. After migration, data gathering for the utilisation of the CRM platform to gain customer insights will be a continuous process. It conventionally requires data integration from various source systems, creation of data models to support integration, and creation and application of metadata definitions (Stone et al., 2017). Data integration refers to combining data that reside at different data sources and requires integration to be of use (Nam, Lee, & Lee, 2019). The source systems may include third-party software, organisational database systems, transactions, and networked touch-points, such as social media or email (Cruz-Jesus, Pinheiro, & Tiago, 2019).

2.2 CRM Data

This section describes the data that is utilised in CRM platforms according to literature. To answer [RQ2] the CRM data types are investigated and their quality is defined. Building on this, literature is reviewed in order to find methods for the measurement of the quality of CRM data, answering [RQ3]. Subsequently, existing DQ challenges and potential solutions are examined ([RQ4]).

2.2.1 CRM Data Types

For this research, data represents the foundation for generating CRM-related information, serving as a key enabler for efficient processes (Otto et al., 2007). According to Zahay et al. (2012) and Missi et al. (2005) there are different types of data within CRM. The different data types are *Customer/Prospect Contact Data*; *Demographic*; *RFM/Transactional Data*; *Psychographic Data*; *Customer Touchpoint Data*; and *Personalisation Data*. They are explained in the following.

- **Customer/Prospect Contact Data:** Customer contact data represents the most basic data in the a CRM platform. It is critical to the operation of CRM platforms and includes customer and prospect information and others involved in the customer journey.
- **Demographic:** Demographic data includes values such as gender, marital stage, income, and ethnicity.
- **RFM/Transactional Data:** Recency, Frequency, Monetary (RFM)/Transactional data includes information such as the customer's last purchase date, frequency of purchase, or customer service activities. The data is required for RFM calculations to analyse customer value, as well as defining customer retention and lifetime value. It is behavioural data and can be classified as predictive.
- **Psychographic Data:** Psychographic data consists of information such as beliefs, opinions, lifestyles, values, or motives of customers. This information can be utilised for the customer relationship processes, as it is an added dimension to predictive modelling and denotes intended behaviour.
- **Customer Touchpoint Data:** Customer touchpoints include information captured via the internet, email click-throughs, service encounters, and telephone calls.
- **Personalisation Data:** Personalisation represents the capability to tailor communications to the individual customer. Its data includes information that is targeted at specific customers.

The types of data mentioned above can be classified into either structured, unstructured, or semi-structured data (Buttle & Maklan, 2019; Nam, Lee, & Lee, 2019). In Table 2.1 an explanation as well as exemplification of the structures can be found.

TABLE 2.1: CRM data structures

Data Structure	Definition	Example
Structured Data	A data structure is provided by a predefined model specifying how the data should be stored	Relational data tables
Semi-structured Data	Data that have a structure with some degree of flexibility	Web page; XML
Unstructured Data	Textual or non-textual data without a pre-defined model; A generic sequence of symbols	Email text; Phone call

2.2.2 CRM Data Quality

The quality of the data can be defined as the fitness to serve a purpose in the context of CRM (Vancauwenbergh, 2019). However, there exist several other approaches towards DQ. It is defined as a multi-faceted construct, consisting of a set of DQ dimensions (Otto et al., 2007; Wang, 1998). Those DQ dimensions are a set of quality

attributes requiring consumer assessment. Dimensions are highly context dependent and their relevancy can vary between organisations and types of data. Hence, the identification of the relevant quality dimensions builds the basis for the assessment of DQ and its possible improvements (Cichy & Rass, 2019). Concluding from section 2.2.1, DQ dimensions need to include structured, semi-structured, and non-structured data structures to define CRM DQ.

Data Quality Dimensions

The identification of the relevant quality dimensions builds the basis for the assessment of DQ and its possible improvements (Cichy & Rass, 2019). There exists an extensive amount of different dimensions and definitions in literature, of which a most common set includes *accessibility*, *accuracy*, *completeness*, *consistency*, *integrity*, and *timeliness* (Batini et al., 2009; Cichy & Rass, 2019; Otto et al., 2007; Pipino, Lee, & Wang, 2002; Wang, 1998; Ballou et al., 1998). This set of dimensions is elaborated on in the following.

- **Accessibility:** Accessibility is defined as the extent to which information is available, or easily and quickly retrievable for those who need it.
- **Accuracy:** The extent to which data are correct, reliable, and certified. Existing DQ management methodologies only consider syntactic accuracy, defined as the closeness of a value, v, to the elements of the corresponding definition domain, D. For DQ management, the only interest is to check whether v is any one of the values in D, or how close it is to value in D. For example, when an organisation obtains the name "John" for a customer, but the actual known name is "Jonas", then the data is not accurate.
- **Completeness:** Completeness refers to the degree in which a given data collection includes data describing the corresponding set of real-world objects, hence whether the data are of sufficient breadth, depth, and scope for the task at hand. Values are missing when they exist in the real world, but are not available in a data collection. It can be missing either because it is unknown while it exists, or because the value does not exist, or because it is unknown whether the value exists. For example, a customer's first and last name are mandatory, but the middle name is optional. The data is complete when the middle name is missing, yet incomplete when the first name is missing.
- **Consistency:** Consistency is the extent to which there exist contradiction in data. Semantic rules specify the correctness of data. For example, data is not consistent when a customer status in the CRM platform is "terminated", yet the invoice status of same customer is "active".
- **Integrity:** The extent to which the data entered in the database have the required format. It is defined as the violation of semantic rules defined over a set of data items. For example, data lack integrity when it does not follow a certain predefined format, such as "dd/mm/yy" for a customer's date of birth.
- **Timeliness:** Timeliness is defined as the extent to which the age of data is appropriate for the task, or the degree to which a value is up-to-date. For example, customer service needs to provide up-to-date information about a product to the customers.

Data Quality Measurement

DQ can either be measured subjectively, for example by asking the data consumer to rate the quality of dimensions, or objectively, by defining metrics consisting of

computations that quantify DQ, giving an indication of the level of DQ (Cichy & Rass, 2019). Metrics associated with each DQ dimension are defined, providing measurements for dimensions with greater (lower) metric values representing a greater (lower) level of DQ and each DQ level being represented by a unique metric value. They are required for two main reasons: (1) the metric values are used to support data-based decision making under uncertainty, and (2) the metric values are used to support an economically oriented management of DQ (Heinrich et al., 2018). In order to make objective assessments, organisations should comply to certain principles and develop metrics specific to their needs (Erceg, Šereš, & Zoranović, 2019; Pipino, Lee, & Wang, 2002). Often, one metric is not sufficient to accurately measure a DQ dimension, and different metrics should be combined to derive a clear picture of the actual DQ (Cichy & Rass, 2019). Once the task of defining the accurate dimensions is completed, the following types of DQ metrics can be formulated (Pipino, Lee, & Wang, 2002):

- **Ratio:** Ratio measures the ratio of desired outcomes to total outcomes. When measuring exceptions, the number of undesirable outcomes is divided by the total outcomes subtracted from 1, where 1 represents most desirable and 0 represents least desirable score. Examples of dimensions that can be evaluated using this type of metric are *completeness*, *consistency*, and *accuracy*.
- **Min or Max operations:** The min or max operation can handle dimensions that require the aggregation of multiple data quality indicators (variables). The minimum (or maximum) value from among the normalized values of the individual data quality indicators is computed. Examples of dimensions that can be evaluated using this type of metrics are *timeliness* and *accessibility*.
- **Weighted Average:** An alternative to the min operation is the weighted average. When an organisation has an understanding of the importance of each variable to the overall evaluation of a dimension, then a weighted average of variables is appropriate. An example of a dimension where this type of metric would be applicable is *believability*. Believability is the extent to which data is regarded as true. Among other variables, it may reflect in a person's assessment of the credibility of the data source, comparison to other accepted standards, and previous experience. When the the importance of each of these variables is known, weighted average can be applicable.

For the dimensions mentioned previously, examples of metrics along with their definitions can be found in Table 2.2.

TABLE 2.2: Dimensions & Metrics

Dimension	Metric	Explanation
Accessibility	Max operations	The maximum value of two terms: 0 or 1 minus the time interval from request by user to delivery to user divided by the time interval from request by user to the point at which data is no longer useful. Delivery refers to when the data is delivered to the user.
Accuracy	Ratio of data to errors	The number of errors in the data over a data set. When the data set size increases or stagnates and the number of errors in the data decreases, the DQ increases.
Consistency	Ratio of data to inconsistencies	The number of inconsistencies in the data over a data set.
Completeness	Ratio of data to blank values	The number of missing values over a data set. An increase in blank values indicates a decrease in DQ.

Continued on next page

Table 2.2 – continued from previous page

Dimension	Metric	Explanation
Integrity	Ratio of data transformation errors to data transformations	Data transformation is the conversion of data into another format. The error rate is the frequency in which the data transformation operation fails.
Timeliness	Max operations	The maximum of one of two terms: 0 and 1 minus the ratio of currency to volatility (Ballou et al., 1998). Currency is defined as the age plus delivery time minus input time and volatility as the length of time data remains valid. Input time to when the data is received by the system and age refers to the age of the data when first received by the system.

Finally, in order to conclude whether the data is of sufficient quality, thresholds to compare the calculated metrics to need to be defined (DAMA, 2017). Whenever a measurement exceeds the predefined threshold, the data can be concluded to be of insufficient quality. This threshold can differ per organisation or moment in time. For example, the *completeness* for the value *last name* in a financial organisation might be of more relevance than for an organisation in the product industry. Each organisation has to adhere to various internal as well as regulatory policies that might define DQ thresholds (DAMA, 2017).

2.2.3 Challenges

The high quality of successful integrated customer data is the foundation of successful CRM projects, while key problems regarding CRM relate to the quality of data and its integration (Missi, Alshawi, & Fitzgerald, 2005). Literature suggests various challenges and solutions regarding CRM DQ, of which an overview can be found in Table 2.3.

Inconsistencies between the ways in which information is held by different units in an organisation cause problems when the data is made organisation-wide and used to analyse customers. Furthermore, data sufficient for operational systems has often proved insufficient for CRM platforms (Reid & Catterall, 2014). Additionally, contemporary CRM platforms make use of Data Mining (DM) approaches to assist in CRM by analysing the large amounts of customer data and find the best CRM strategies (Natchiar & Baulkani, 2014). For DM techniques to work effectively, high DQ is crucial. However, big data for CRM yields several DM problems for traditional DM approaches due to its volume, velocity, and variety, which makes the management of CRM DQ increasingly complex (Batini et al., 2009). The problems include heterogeneity, dimensionality, serious anomalies on data, unbalanced classification, data encryption, etc. (Tu & Yang, 2013).

TABLE 2.3: Identified challenges & solutions for CRM data quality

ID	Challenge	Solution
C1	Lack of agreement on a standard set of DQ dimensions	The creation of a ranking on allowable expected DQ dimensions in CRM platforms, which depends on the specific data elements (Sharma, Goyal, & Mittal, 2010). Different records in a CRM platform may require different levels of DQ defining optimal quality. Even et al. (2010) propose a model that allows different levels of DQ to be set for different records in a CRM platform.

Continued on next page

Table 2.3 – continued from previous page

ID	Challenge	Solution
C2	Inconsistency in data entry an storage	Creation of a framework for data standards and rules, as well as a process for the verification of data (Tu & Yang, 2013; Petrović, 2020); Analysis of whether it is more beneficial to focus on process redesign or simplification rather than on the integration of an existing process (Foss et al., 2002). For example, the way of collecting customer data can lead to increased customer DQ in CRM platforms (Zahay, Krishen, & Peltier, 2012). E.g., adjusting the capture frequency of data in CRM platforms.
C3	Data defects, such as incorrect or missing data, unique values or syntax violations, or outdated values	Improvement of the data entry processes by including data verification processes addresses data completeness (Petrović, 2020); Automation of data collection; Continuous inspection, correction and cleaning (Even, Shankaranarayanan, & Berger, 2010); utilisation of error detection tools (Missi, Alshawi, & Fitzgerald, 2005).
C4	Condition of existing customer databases	To improve DQ to the level required for the new CRM platform, the data needs to be cleaned and managed before being transferred to or integrated with the databases that are utilised by the new CRM solution. This can be done by using a data cleanup tool (Goldenberg, 2008; Reid & Catterall, 2014).
C5	Inadequate integration of data sources	Usage of specialized tools or data models for the integration of data of different formats (Petrović, 2020; Stone et al., 2017).
C6	DQ deteriorating over time	(Automatic) Updates and linkage of data throughout the entire CRM platform (Petrović, 2020).
C7	High costs for maintaining DQ	Cost reduction can be achieved by addressing key quality problems from the start of the CRM implementation project (Even, Shankaranarayanan, & Berger, 2010).
C8	Big data	Big data creates problems for conventional DM techniques used in CRM, because of its volume, velocity, and different data structures (Nam, Lee, & Lee, 2019). The proposed solution is big data analytics, for which cloud computing provides the infrastructure (Balachandran & Prasad, 2017). The SaaS firm should have a standardized framework for big data analysis in place, which could produce satisfactory results for the organisation's CRM data (Tu & Yang, 2013).

Concluding from the identified DQ challenges and potential solutions, the DQ management framework for CRM should include practices for the definition of data (C1), definition of standards and rules (C2), DQ assessment practices (C2, C3, C4, C8), development of a DQ improvement strategy (C2, C3, C4), integration and migration practices (C4, C5), and implementation of DQ management practices continuous and from the start of the CRM solution (C6, C7).

2.3 Data Quality Management

This section reviews existing literature about DQ management methods to provide an answer to [RQ5] in the context of CRM platforms. Data management concerns defining the data architecture including organisation-wide data models and data modelling standards, data management processes, and data management systems design and operations. DQ management refers to quality-oriented data management, where the management focuses on the the collection, organisation, storage,

processing, and presentation of data of high quality (Otto et al., 2007). In the following sections, existing DQ management methods are explained, after which the applicability to CRM is elaborated on.

2.3.1 Existing Data Quality Management Methods - A Review

Existing scientific literature on DQ management methods appears to be scarce (Wijnhoven et al., 2007) and rather focuses on DQ by itself or solely its assessment. However, for CRM to be successful, it is important to create a comprehensive DQ management strategy at the beginning of a CRM implementation (Reid & Catterall, 2014), where DQ management is a continuous process. Spruit and Pietzka (2015) found that the management of DQ is optimized when DQ assessment is performed regularly, the costs of poor DQ can be expressed in monetary as well as non-monetary terms (the business impact is determined (Spruit & Linden, 2019)), the organisation recognises the root causes of poor DQ, and the organisation establishes improvement practices when DQ is found to be poor. Literature demonstrates that the maintenance of high-quality data in general requires a method whereby companies cyclically audit and clean the data, as well as implement compliance measures for their data repositories (Marsh, 2005). Hence, to be of relevance for this research, the methods should contain the main steps of DQ management, namely *DQ definition*, *DQ assessment*, and *DQ improvement* (Cichy & Rass, 2019), executed in a cyclical manner. Furthermore, the methods cannot lack generality, as they should be relevant for data within different fields and industries. To determine the applicability of the DQ management methods to CRM, an additional set of criteria considering the methods' common points and the factors that influence the solution to challenges as explained in section 2.2.3 is formulated (Francisco et al., 2017; Spruit & Pietzka, 2015). The criteria are listed below, along with the activities that should be identified within the method and the related solution(s).

- **Upfront considerations:** Reconstruct the organisational environment. Define a data standardization that is organisation-wide. This should include structured, semi-structured, and unstructured data.
- **DQ assessment:** DQ is defined regarding the requirements is different stakeholders in terms of DQ dimensions. It is measured either objectively or subjectively for every group of data. DQ should be assessed frequently according to the standard set of DQ dimensions to identify DQ problems, considering all possible integrations and data sources.
- **Impact Analysis:** Analyze the business impacts of poor DQ in a monetary as well as non-monetary way.
- **Root Cause Detection:** Detect the root causes of poor DQ. Be aware of the whereabouts of the existing root causes.
- **DQ improvement:** Activities for selecting data improvement techniques should take place. Activities for data transformation, such as data cleaning and refining the root cause of the DQ problem should take place. This should be data-driven as well as process-driven.

An overview of some of the commonly known and reviewed methods found in literature that meet the majority or excel in some of previously mentioned criteria can be found in Table 2.4 (Batini et al., 2009; Cichy & Rass, 2019). A more detailed description along with the phases of the individual methods can be found in Appendix A.

TABLE 2.4: Data quality management methods

Acronym	Name	Reference
AIMQ	A methodology for information quality assessment	(Lee et al., 2002)
CDQ	Comprehensive methodology for Data Quality management	(Batini et al., 2006)
COLDQ	Cost-effect of Low Data Quality (Loshin Methodology)	(Loshin, 2004)
DQA	Data Quality Assessment	(Pipino, Lee, & Wang, 2002)
DQAF	Data Quality Assessment Framework	(Batini et al., 2015)
DQPA	A Data Quality Practical Approach	(Cichy & Rass, 2019; Pilar Angeles & García-Ugalde, 2009)
HDQM	A Data Quality Methodology for Heterogeneous Data	(Batini et al., 2011)
HIQM	Hybrid Information Quality Management	(Cappiello, Ficiaro, & Pernici, 2006)
OODA DQ	The Observe-Orient-Decide-Act Methodology for Data Quality	(Sundararaman & Venkatesan, 2017)
TBDQ	Task-Based Data Quality Method	(Vaziri, Mohsenzadeh, & Habibi, 2016)
TDQM	Total Data Quality Management	(Wang, 1998)
TIQM	Total Information Quality Management	(English, 1999)

The distinct methods adopt different names and details for the individual steps. For example, TDQM, DQA, and OODA DQ do not provide formal steps for the assessment process. For TDQM, assessment consists of defining metrics and implementing those. DQA simply suggests that to assess DQ, defined metrics need to be applied. OODA DQ suggests that assessment is related to the first phase of the method, namely *Observe*. For AIMQ and CDQ, assessment consists of problem identification through interviews and quantitative evaluation of DQ issues. HIQM provides a unique step called *Warning*, where a components diagnoser, feedback modules, message generator, warning log database, warning analyzer, warning/recovery database, and a real-time recovery module are comprised. HDQM starts by ranking resources to establish the feasibility and risk for the improvement phase, after which a quantitative measurement of DQ takes place. The assessment of TBDQ consists of defining the goals and scope of DQ, and assigning weights to DQ issues by means of a comparison matrix. TBQM suggests questionnaires for subjective assessment, and a simple ratio metric for objective assessment. DQPA consists of several steps for DQ assessment, including identification of DQ properties, analyzing existing metrics, describing methods for representing and assessing DQ indicators, storing metadata containing quality scores of data sources.

Despite those differences, a general case with a set of basic steps composed of three phases can be extracted: *state reconstruction*, *assessment (including DQ definition)*, and *improvement* (Batini et al., 2009). The different phases are elaborated on in the following paragraphs, along with an explanation of the main differences between the existing DQ management methods.

State Reconstruction

State reconstruction aims to collect the contextual information on the organisation, data, and related processes. This phase is considered optional if the assessment phase can be based on existing documentation, which is assumed by the majority of the DQ management methods. Therefore, it is often not explicitly mentioned within the phases of the methods.

Assessment

The assessment of DQ is a critical part of DQ management. Assessment is aimed at measuring the quality of data using defined DQ dimensions and metrics. Metadata plays a key role in the assessment phase, as this stores complementary information for, among other things, DQ. It provides the information required for understanding and evaluating the data. For the reviewed DQ management methods, the overall description of the assessment phase and the individual steps differ significantly in regards to the degree of detail (Cichy & Rass, 2019). In general, the assessment phase consists of the following set of basic steps (Batini et al., 2009):

1. *Data analysis* to reach a complete understanding of data and related architectural and management rules.
2. *DQ requirements analysis* to identify quality issues and set new quality targets.
3. *Identification of critical areas* to select the most relevant databases and data processes to be assessed quantitatively. The comprises an impact analysis of poor DQ in monetary as well as non-monetary terms (Spruit & Pietzka, 2015; Spruit & Linden, 2019; Heinrich et al., 2018). Spruit and van der Linden (2019) propose eleven business impacts of poor DQ to be included in the business impact analysis.
4. *Process modeling* to provide the processes that are producing or updating data.
5. *Measurement of quality* to select the quality dimensions affected by the quality issues identified in previous steps and define corresponding metrics.

To benchmark the reviewed methods regarding the assessment phase of DQ management, the considered data types, dimensions, and measurement techniques/ strategies per method are reviewed. The majority of the reviewed methods mainly considers structured data. TBDQ and DQPA only consider structured data, but envision to extend their models to an investigation of DQ to other data types. HDQM considers structured, semi-structured as well as unstructured data by translating the different types of data resulting from heterogeneous resources into a common, conceptual representation. The measurement techniques in AIMQ may also apply to both structured and unstructured data, as it considers semi-structuctured data implicitly. For TDQM and CDQ semistructured data is considered as well. TIQM and COLDQ consider semi-structured data implicitly.

The reviewed methods recognize different dimensions depending on their focus area. Apart from the differences in dimension selection, the definition of the individual dimensions can also vary per method. In Table 2.5, the dimensions used for DQ assessment that are explicitly mentioned per reviewed DQ management method can be found, along with whether the method supports or suggests the extension to further dimensions. For TIQM, DQA, and TBDQ it is particularly suggested to identify the dimensions that are most important for the specific organisation. To do so, TIQM suggests a survey for DQ expectations of current as well as prospective data consumers. In DQPA the DQ dimensions are selected through expert user judgements and depend on the type of information system and relevant data. For CDQ, the DQ dimensions that are suggested depend on the data structure. The various adoptions of DQ dimensions within the distinct methods confirms the gap regarding an effective standardization of dimensions as mentioned in section 2.2.3.

TABLE 2.5: Data quality management methods dimensions

Method	Dimensions	Extension
AIMQ	Accessibility, Appropriateness, Believability, Completeness, Concise/Consistent representation, Ease of operation, Freedom from errors, Interpretability, Objectivity, Relevancy, Reputation, Security, Timeliness, Understandability	No
CDQ	Schema: Correctness with respect to the model, Correctness with respect to Requirements, Completeness, Pertinence, Readability, Normalization, Data: Syntactic/Semantic Accuracy, Semantic Accuracy, Completeness, Consistency, Currency, Timeliness, Volatility, Completability, Reputation, Accessibility, Cost	Yes
COLDQ	Schema: Clarity of definition, Comprehensiveness, Flexibility, Robustness, Essentialness, Attribute granularity, Precision of domains, Homogeneity, Identifiability, Obtainability, Relevance, Simplicity/Complexity, Semantic consistency, Syntactic consistency - Data: Accuracy, Null Values, Completeness, Consistency, Currency, Timeliness, Agreement of Usage, Stewardship, Ubiquity, Presentation: Appropriateness, Correct Interpretation, Flexibility, Format precision, Portability, Consistency, Use of storage, Information policy: Accessibility, Metadata, Privacy, Security, Redundancy, Cost	No
DQA	Accessibility, Appropriate amount of data, Believability, Completeness, Freedom from errors, Consistency, Concise Representation, Relevance, Ease of manipulation, Interpretability, Objectivity, Reputation, Security, Timeliness, Understandability, Value added	Yes
DQAF	Completeness, Timeliness, Validity, Consistency, Integrity	No
DQPA	Accuracy, Completeness, Consistency, Currency, Timeliness, Uniqueness, Volatility	No
HDQM	Accuracy, Currency	Yes
HIQM	Accuracy, Completeness, Consistency, Timeliness	Yes
OODA DQ	Speed, Volume	No
TBDQ	Accuracy, Completeness, Consistency, Timeliness	Yes
TDQM	Accessibility, Appropriateness, Believability, Completeness, Concise/Consistent representation, Ease of manipulation, Value added, Free of error, Interpretability, Objectivity, Relevance, Reputation, Security, Timeliness, Understandability	No
TIQM	Inherent dimensions: Definition conformance (consistency), Completeness, Business rules conformance, Accuracy (to surrogate source), Accuracy (to reality), Precision, Nonduplication, Equivalence of redundant data, Concurrency of redundant data, Pragmatic dimensions: accessibility, timeliness, contextual clarity, Derivation integrity, Usability, Rightness (fact completeness), cost	No

The DQ measurement strategies and techniques differ between the distinct methods. As mentioned in section 2.2.2, the dimensions can be measured either subjectively or objectively. Most methods rely on objective metrics, some methods suggest a combination of objective measures and subjective assessments, and AIMQ relies solely on subjective measurements. DQAF provides a set of objective DQ metrics to choose from. TDQM provides some common metrics to measure DQ dimensions. Additionally, TDQM takes into account that certain business rules need to be considered when assessing DQ. OODA DQ does not suggest specific metrics, and rather suggests to choose between a wide range of metrics derived from literature. An overview of explicitly mentioned metrics per method can be found in Table 2.6.

TABLE 2.6: Data quality management methods metrics

Method	Measurement
AIMQ	Subjective measurement through a survey questionnaire
CDQ	User interviews; Definition of metrics for the dimensions accuracy and currency
COLDQ	Consumer surveys; Definition of a variety of metrics
DQA	Stakeholder expectations; Definition of quantitative metrics (simple ratio, min or max operations, weighted average)
DQAF	Definition of a set of metrics for different for different types of measurement
DQPA	Definition of primary data sources; Derived metrics
HDQM	Definition of metrics for the dimensions accuracy and currency
HIQM	Objective assessment through a measurement algorithm suggested
OODA DQ	-
TBDQ	Survey questionnaire; Simple ratio
TDQM	Consideration of business rules; Definition of metrics
TIQM	User expectations; Definition of metrics

Improvement

This phase defines the steps and strategies that need to be undertaken for reaching new DQ targets. Organisations need to consider different tools and techniques, while taking into account the costs. The improvement phase commonly consists of the following set of basic steps (Batini et al., 2009):

1. *Evaluation of costs* to estimate the direct and indirect costs of DQ. Direct costs are the costs of assessment and improvement activities. Indirect costs are the process costs caused by data errors and opportunity costs due to lost revenues. In the context of economically oriented management of DQ, DQ improvement measures should be applied if and only if the benefits (due to higher data

quality) outweigh the associated costs (Heinrich et al., 2018). Batini et al. 2007 provides a detailed and complete classification of costs.

2. *Assignment of process responsibilities* to identify the process owners and define their responsibilities on data production and management activities.
3. *Assignment of data responsibility* to identify the data owners and define their data management responsibilities.
4. *Identification of causes of errors* to identify the root causes of DQ issues.
5. *Selection of strategies and techniques* to identify the data improvement strategies and corresponding techniques that comply with the context of CRM and the specific organisation. The improvement strategies can be either *process-driven*, meaning the quality of data is improved by redesigning processes that produce or modify data, or *data-driven*, meaning the quality of data is improved by directly modifying the value of data. Examples of techniques applied by data-driven strategies are *acquisition of new data*, *standardization of data*, *record linkage*, and *data integration*. Examples of techniques applied by process-driven strategies are *process control* to insert checks in the data production process, and *process redesign* to remove the causes of poor DQ and introduce new activities that produce higher DQ.
6. *Design of data improvement solutions* to select the most effective and efficient strategy and related set of techniques and tools to improve DQ.
7. *Process control* to define checkpoints processes producing data, to monitor DQ during execution.
8. *Improvement management* to define organisational rules for DQ.
9. *Improvement monitoring* to establish periodic monitoring activities.

TDQM, DQA, and DQAF emphasize the root causes of DQ issues. HIQM suggests data-oriented as well as process-oriented improvement steps, including possible changes at a strategic level. TDQM, TIQM, COLDQ, and CDQ also suggest process-driven strategies next to data-driven strategies. DQAF emphasisez the comparison between DQ assessment and DQ expectations. There are also differences in cost considerations between the methods. AIMQ, DQA, HIQM, and OODA DQ do not consider costs explicitly. Most of the remaining methods include a cost-benefit analysis, except for DQPA and TDQM. A cost-benefit analysis is a process in which benefits and costs of a project are compared systematically and analytically in order to assess its value (Cichy & Rass, 2019). A majority of the methods emphasizes that there should be a decision process in place to decide where improvements should be made. The different strategies that support improvement decision processes that are explicitly mentioned in the distinct methods can be found in Table 2.7.

TABLE 2.7: Data quality management methods decision strategies for data quality improvements

Method	Strategy
AIMQ	Based in DQ categorization and gap analysis
CDQ	Based on the state reconstruction phase
COLDQ	DQ scoreboard to identify the best opportunities for improvement
DQAF	Definition of directives for DQ strategy

Continued on next page

Table 2.7 – continued from previous page

Method	Strategy
DQPA	Ranking of business questions and data sources to determine impact and required improvements
HDQM	ReSource-Improvement-Activity Matrix
TBDQ	Prioritization of data units based on weight and measurement score; award system
TDQM	Information manufacturing analysis Matrix; Integer programming model

2.3.2 Data Quality Management within a CRM Delivery & Consultancy Project

Concluding from literature, the utilisation of a mix of the existing DQ management methods for the context of CRM should be applied, using the best of them (Francisco et al., 2017; Cichy & Rass, 2019; Otto et al., 2007; Batini et al., 2009). This means that a framework for management of DQ in CRM D&C projects can be generalized into a combination of the CRM D&C practices of data migration and integration, and the mix of reviewed DQ management methods. As aforementioned, the migration and integration practices of CRM D&C projects generally consist of an ETL process. When combined with the DQ management practices of the reviewed DQ management methods, Transform of ETL corresponds to the Assessment and Improvement phases of DQ management, as this comprises the transformation (read: improvement) of data where this is required for successful delivery of the new CRM platform. Furthermore, in order to facilitate the aforementioned variation in data types that are required for a CRM platform, the definition of DQ in subjective as well as objective manners should be taken into account. Hence, the definition of DQ dimensions and accompanying metrics that are required will vary per CRM D&C project, thus needs to be an activity in each CRM D&C project. Below list of activities are concluded from literature to be included in a CRM-DQMF:

- Upfront considerations to define the CRM D&C and reconstruct the organisational environment.
- Data migration/integration to incorporate CRM D&C projects practices.
- Continuous DQ assessment including a business impact analysis to identify critical data elements, to define and measure DQ. Tho achieve this, assigned data owners are responsible for the data.
- DQ improvement including the detection of root causes of DQ issues, as well as a decision-making process based on a cost-benefit analysis of improvement strategies.

Chapter 3

Investigation of Data Quality Management in CRM Delivery & Consultancy

Within this Chapter, the case study of this research is elaborated on. First, the aim and relevance of the case study is explained. Second, the sampling results of the documentation analysis and expert interviews are provided, followed by the triangulated results of the entire case study by means of the thematic analysis.

3.1 Aim & Relevance

The aim of the case study is to scrutinize the current practice of DQ management in CRM D&C projects, as well as DQ management requirements for CRM D&C projects. It extends the current literature on the topic of DQ management, and projects its insights onto the context of CRM D&C projects. The results of the case study serve as critical input for the initial design of the CRM-DQMF. Since there does not exist extensive literature on DQ management in the context of CRM D&C, the case study is relevant within the current state of the DQ management and CRM D&C literature.

3.2 Sampling Results

This section contains the sampling results that are utilised for the case study as part of this research. First, the results for the documentation analysis are shown. Second, the participants of the expert interviews.

3.2.1 Documentation Analysis

In Table 3.1 the documentation that is consulted for this case study can be found. A total of 16 different documents are identified to be of relevance to this research. The documents include documents that are used for training purposes of employees, documentation of the CRM platform (D365) of this case study, and documents containing information on the execution of specific CRM D&C projects.

TABLE 3.1: Documents acquired for documentation analysis

ID	Purpose	Author	Doc. Type	Year	Version
Continued on next page					

Table 3.1 – continued from previous page

ID	Purpose	Author	Doc. Type	Year	Version
D1	D365 insights	Microsoft	PDF	2020	-
D2	Data Migration guidelines	Director technology leadership at Avanade	ADF	2020	9
D3	Dynamics CRM (on-premises) to D365 migration guidelines	Microsoft	DOC	2020	-
D4	D365 Adoption guidelines	Microsoft	DOC	2020	-
D5	Microsoft Common Data Model	Microsoft	DOC	2020	-
D6	Report on the solution design #1 for a utility client of Avanade	Project employees at the client	DOCX	2020	-
D7	Report on the solution design #2 for a utility client of Avanade	Project employees at the client	DOCX	2020	2
D8	Data flows explanation within a solution for a utility client at Avanade	Project employees at the client	DOCX	2020	-
D9	Report on solution design #2 for utility client of Avanade	Project employees at the client	DOCX	2020	2
D10	Functional design datamart for a utility client at Avanade	Project employees at the client	DOCX	2020	2
D11	Education in D365 Marketing	Senior consultant at Avanade	Training	2021	1
D12	Global policy in data management	IS Program Management at Avanade	PDF	2021	8
D13	Dedicated system for data quality checks in multiple systems	Accenture	PDF	2011	-
D14	Data Management Roadmap Proposal	Accenture	PowerPoint	2021	-
D15	Data Quality Analysis strategy at a product client at Avanade	Analytics at Avanade	PowerPoint	2019	1.5
D16	Design principles in a CRM project of an NGO client at Avanade	Senior consultant at Avanade	Powerpoint	2021	1

3.2.2 Expert Interviews

In Table 3.2 the participants of the expert interviews can be found. A total of 14 experts have participated in the case study. The experts provide knowledge from different areas (CRM and Analytics), and different industries on varying levels of experience. The majority of the participant, 12 out of 14, work at the CRM department at Avanade, focussing on solely CRM projects. There was chosen to include two participant of the Analytics department, as those participant have worked together with the CRM department on CRM projects from a data analytics point of view. Those participants might have a more in-depth knowledge of DQ related challenges and solutions.

TABLE 3.2: Participants Expert Interviews

ID	Role	Department	Industry	Description	Experience
P1	Consultant	CRM	Finance, Utility	Solution developer leading a technical team	4
P2	Senior Analyst (Functional)	CRM	Utility	Business related consulting in varying CRM projects	2
P3	Senior Director	CRM	Finance, Utility, Product, NGO	Business responsible for everything concerning CRM, ERP, and digital marketing	14
P4	Group Manager	CRM	Finance, Product	Business related consulting in CRM with a focus on sales management	15
P5	Senior Consultant (Functional)	CRM, Marketing	Finance, Utility, Product, NGO	Business related solution architect of marketing/communication in CRM	13
P6	Senior Consultant (Functional)	CRM	Finance, Utility, Product, NGO	Business related consulting in CRM projects	5
P7	Senior Analyst	CRM	Finance	Consulting in CRM projects	4
P8	Manager	CRM	Finance, Utility, Product	Project- or delivery manager (technical), solution architect	9
P9	Consultant (Functional)	CRM	Finance, Utility, Product	Consulting in CRM projects	4
P10	Consultant (functional)	CRM	Utility	Business related consulting in CRM projects	6

Continued on next page

Table 3.2 – continued from previous page

ID	Role	Department	Industry	Description	Experience
P11	Analyst (Functional)	CRM	Finance, NGO	Business related consulting in CRM projects	2
P12	Senior Consultant (Functional)	CRM	Finance	Business related consulting in CRM projects	7
P13	Director	Analytics	Finance, Utility, Product	Manage projects from start till end as project lead or architect, analyze projects on reusability, focus on Power BI	14
P14	Senior Consultant	Analytics	Finance, Product	Consulting in analytics projects, sometimes combined with CRM projects	7

3.3 Triangulated Results

The triangulated results of the case study, which include the results of the expert interviews as well as the results of the documentation analysis, are expressed by means of a thematic analysis. Three main themes were extracted, namely *Modularity*, *CRM D&C*, and *DQ Management*. In Appendix B.4, the code book of the thematic analysis can be found. In the following, the extracted themes are elaborated on by means of the extracted codes that belong to the theme. Finally, a summary of the results is provided, which concludes the main findings of the case study.

3.3.1 Modularity

The first theme that is extracted is modularity, which describes the uniqueness of every CRM D&C project, and therefore the need for uniqueness in DQ management application in CRM D&C projects. The execution of DQ management depends on the unique situation of every project and client. First of all, every project has its own client context. Second, every client has its own level of DQ management expertise. The separate codes are elaborated on in following paragraphs.

Client Context: One of the participants (P3) explained that the challenge a CRM D&C team faces is that there already exists a context, namely the client's context, with its own constraints. The task of the CRM D&C team is to deal with those constraints and support the client in running through their decision processes successfully. To do so, the correct information needs to be gathered, interpreted correctly, and presented to the client to help them make the correct decisions. Some participants also call the client context a *blueprint*, which is defined by P3 as follows: "The blueprints create the framework for the CRM solution, based on budget, goal, and scope. Blueprints contain scope, the relevant dataset, functional plan. It's implicit, not explicit. The scope and framework are given by the client, in which the consultant offers a solution". Additionally, business processes are reconstructed as

stated by P2: "together with the client we define and design business process flows", and from documentation is derived that data policies and data standards are defined for the specific project. The constraints include the business goals of the client, which describe what the client wants to achieve with this project. Additionally, it includes the scope of the project, which is the method that is used to achieve the business goals. Finally, it includes the budget of the client for the CRM D&C project.

DQ Management Expertise Level: Various experts indicated that the quality of data, and therefore the need for management of DQ within a CRM D&C project along with the role of the CRM D&C team, depends on the expertise of the client. On the one hand, the client might not have any knowledge on their own data, nor its quality, which might indicate that the data is not of sufficient quality for a CRM solution and the client does not have sufficient DQ management expertise in-house, meaning the expertise of the CRM D&C team is required to establish DQ management. On the other, the client might already be in control of its data (and quality) across the organisation, which means there is no need for the CRM D&C team to conduct or advise on any DQ management practices. One participant (P3) mentioned: "In the ideal case, organisations already have an authority in place that takes care of data quality matters. However, this varies per organisation and industry. The interference of Avanade depends on the arrangements with the client". Additionally, from the case study various DQ Key Performance Indicator (KPI)'s are mentioned, of which some examples can be found in Table 3.3.

These insights indicate that there is need for the determination of the client's DQ management maturity level within the specific CRM D&C project, which determines the extent in which DQ management will be applied and by whom.

TABLE 3.3: KPI's of Data Quality

KPI	Description
The number of data elements with a definition and standardization	This gives an indication of the number of terms with structural quality measurements
Top five data quality issues	The top quality issues give an indication of the most important data quality areas of improvement
Outcome of the tackled data quality issues	The outcome per quality issue demonstrates the value of Data Management

3.3.2 CRM Delivery & Consultancy

The second theme that is extracted is CRM D&C, which describes what a CRM D&C project entails. A CRM D&C team supports its client in reaching its business goals within the scope of the specific CRM D&C project. Generally, this comprises the migration and/or integration of data in a CRM solution.

Business Requirements: The case study demonstrates that the continuous definition and refinement of the requirements of the client's business are essential to the establishment of a successful CRM solution. Without business requirements, business goals cannot be achieved. Business requirements are described by business rules that provide the answer to *"How can the business requirements be met?"*. Business

rules should be aligned with DQ dimensions, which provide the definition of DQ together with a DQ threshold. In Table 3.4 the differences between the terms as explained according to the case study can be found, along with examples.

TABLE 3.4: Definitions

Term	Description	Example
Business Goal	The business goal of the CRM project	Improve the marketing strategy by implementing this in a CRM system
Business Requirement	What needs to be done to achieve the goal?	Send an email to a certain sample of end-customers once a month
Business Rule	How can the business requirements be met? Description of how the data should exist in order to be useful and usable for the business requirements	Two contacts cannot have the same email address; email addresses should contain an "@"
DQ Requirement	Requirement for a data element expressed in DQ dimensions and DQ thresholds	Uniqueness (The data needs to be unique for at least 98% of the records); Conformity (The data needs adhere to a specific format for at least 90% of the records)

Data Mapping: For the migration and/or integration of data, the case study indicates that a data mapping is required. For example, P4 explained that "Business requirements are defined in order to decide which data, which filters, are required for the solution. From this, a data mapping is concluded to load the data to the solution correctly". P2 explained how a migration or integration is established: "We request the required data from the client, and what the client does with this data. Then we show what the CRM solution has to offer". Hence, as concluded form the case study, data mapping is done to map two data models so data can be migrated and/or integrated correctly. This is done by creating a profile of the legacy system, and map this to the profile of the new CRM platform. Once this is done correctly, a migration/integration plan needs to be established as stated by P4: "Integration and migration need to be managed, it requires a plan".

Agility: Statements such as "We act on behalf of the client" (P3) were common within the case study, following the client's wishes and needs. The client decides what will be included in the CRM solution, and every client has different needs. Therefore, most CRM D&C projects within the case study have an agile and customer-centric approach. The agility of an organisation describes the ability to respond faster to changes coming from customers and market through continuous delivery in smaller iterations (Tseng & Lin, 2011). Customer-centricity describes the organisation's ability to develop the CRM solution according to the preferences and needs of the specific customer (Liang & Tanniru, 2006). From this case study, these abilities are found to be main drivers for the D&C of CRM solutions. In the CRM D&C

projects within the case study, this manifests in continuous definition and refinement of business requirements, business rules, and data mapping of the CRM solution.

3.3.3 Data Quality Management

The third theme is DQ Management, which describes what DQ management entails according to the case study results.

Iteration: The case study indicates that DQ management should adopt an iterative approach where all stakeholders are engaged frequently from the start of the project. In the case study this is also indicated as monitoring, which indicates the continuity of DQ assessment and improvement to monitor the quality of data. To give an example, P13 explained the data life cycle, where three steps can be extinguished: (1) Creation of data and (2) management of data (such as modifying, updating, or moving) where all sort of things can go wrong regarding DQ. The last step is (3) the destruction of the data. Up untill the destruction of the data, DQ should be managed. P12 mentioned: "I think data quality should be measured frequently in any case", and participants agreed that ideally this is done from the start of the project, as same participant said: "Resolving data quality issues before migration would make life a lot easier". However, all participants experienced that in practice, the assessment and improvement of DQ is not done proactively. P1 mentioned on DQ requirements: "This is something that needs to be taken into account from the beginning of a project. However, to my experience, this is something that is handled later on".

DQ Definition: The participants of the expert interviews as well as the documentation provided various definitions of DQ. The most common definitions were *consistency* and *completeness*, respectively explained as the need for the data to be synchronised across systems without contradictions, and the need for the data to be without missing values. Other definitions of DQ that are mentioned by experts as well as found in documentation are *serves its purpose* (and is not without purpose), *availability* (the data needs to be available at the right time to the right users through security and access management), *uniqueness* (the data needs to be unique), *awareness* (stakeholders need to be aware of the existence of all data), and *timeliness* (data needs to be up-to-date). One participant (P13) mentioned on the meaning of DQ: "For some use cases, when 10 or 20% of the data is missing, the solution is still working. For other other use cases with other goals, when one decimal is incorrect, major problems will arise", which indicates the use of DQ thresholds for the definition of DQ. This variety in DQ definitions confirms the need for including DQ definition activities in the CRM-DQMF, because of its situational dependence.

Critical Element: In the case study, there is spoken of a *Top Down* or *Demand Driven* approach, where the focus of DQ management is on elements that are relevant to the business. A critical element represents a data element that is of utmost importance to the success of the CRM D&C project solution. It has to comply with defined DQ requirements.

Data Roles: From the case study, several data-related roles were extracted, which are generally realized to steer DQ initiatives. According to most participants, DQ is managed more efficiently whenever someone is responsible for the data, and some

participants even state that without responsible people, DQ cannot be managed at all. For example, P13 mentions: "A DQ KPI on its own does not do anything, someone should own the data. Without ownership, the quality of data will not improve", and P6 says "We always try to assign someone who is responsible for the data at the client's side". To give some examples of data roles, so-called *data champions* were identified to overcome barriers, since they have knowledge of the data and its processes. They build the bridge between the CRM business and the data analysis communities, which is useful for DQ management practices. In terms of responsibility for specific data, *data stewards* were mentioned. Data stewards assure that data are accurate, with high quality, and they establish and monitor sharing of data. When it comes to responsibility of data and its quality, the term *Chief Digital/Data Officer* was mentioned. Concluding from these insights, the assignment of data roles should be included in the CRM-DQMF to assure optimal management of DQ.

Impact Evaluation: According to the case study, in practice, generally only the costs of the implementation of a CRM solution are taken into account for the CRM D&C project proposal. P5 mentioned: "I think that only with bigger organisation, where more money is available, taking into account the costs of poor data quality for decision making might be part of the culture". However, participants explained that in ideal projects also the costs of poor DQ are evaluated. P14 mentions: "An impact analysis should be included in the decision making of what data quality means. However, this is only included occasionally for bigger projects. Often it is more generic". Documentation also showed that the costs that should be taken into account do not solely include monetary costs, but also other business impacts, such as customer satisfaction.

Assessment: In the case study, DQ assessment is described in various terms. It is included in the Transform phase of the ETL process that is used in migration practices. P9 said: "For migration, a three-step process is used, namely the ETL process". Another term that is used is *reporting on DQ issues* to concerned stakeholders, which requires the definition and measurement of DQ. P14 told that when insights on data are provided, ideally, reports are created based on DQ KPI's to provide insights of the quality of that data.

Root Cause: The case study confirms the importance of the detection of the root cause of DQ issues. This is indicated in terms as *detection of improvement location*, meaning DQ issues should be resolved at the source, or a control called *prevention*, which describes that preventing DQ issues relies on a broad spectrum of processes and best practices. P9 states that "The causes of data quality issues are crucial", and P14 explains: "Ideally, a feedbackloop is created where your system of record indicates whether data input processes are performed correctly to ensure the data quality issue is fixed at the source. The places where data is created, should contain the most checks".

Improvement: The improvement of DQ issues is also mentioned as a control called *resolution*. Different improvement strategies came up, both data-driven and process-driven. P6 mentions: "Sometimes we update all data to ensure it is transformed into correct data. What also happens is that we decide that current data cannot be saved anymore (it is too expensive), and only data created from now on will be correct

through an improvement in the process". This indicates the use of a cost-benefit analysis for the development of an improvement strategy. Some other examples for improvement strategies from the case study are: fuzzy matching, where data from different sources is matched to build a complete view of the data; removing nulls; or enhancing the data by using third party sources (e.g. demographic datasets).

3.3.4 Summary of Results

From the case study can be concluded that the level of DQ management in contemporary CRM D&C projects is lacking, as there is no mutual awareness of the importance of DQ management in clients or CRM D&C employees, nor are there best practices in place to perform DQ management integrated in CRM D&C projects. The case study showed that there is need for DQ management, as challenges experienced by participants in CRM D&C projects and found in documentation are largely related to poor DQ. Participants confirmed that, ideally, to provide a complete and accurate as possible solution and advice as a CRM D&C team to the client, DQ management needs to be taken into account in every CRM D&C project. One participant (P12) mentioned "A better integration of DQ management in the project process would prevent many DQ problems. As a project team we would like some practical advice on in which phases of the project we need to look at DQ, and how this should be done".

The case study contributes to the findings of the literature review in two different ways: (1) By confirming literature findings on DQ management and CRM D&C projects, and (2) by providing new insights on the current level of DQ management in CRM D&C projects, the level of necessity for the topic, and the manner on which DQ management could be integrated in CRM D&C projects. The confirmations of literature findings are summarized in the following:

- As stated in section 1.1, CRM platforms are becoming increasingly interconnected and complex as a result of growing variety and needs of customers (Stone et al., 2017), who expect the CRM platform to be tailored for their organisation specifically, with limited effort, and deployed promptly (Cricelli et al., 2019). The insights as found in the case study regarding customer-centric and agile project approaches enable D&C teams to deal with this, and thus correspond to these literature findings.
- Upfront considerations as part of the criteria for DQ management in CRM D&C projects concluded from literature as listed in section 2.3.2 is included in the case study in terms of the definition of a client context, also called a blueprint. This includes a business case with scope, business goals, and budget, as well as a state reconstruction of the concerned data, the business processes, data policies, and data standards.
- The assignment of data roles to define DQ management responsibilities as found to be a relevant part of any DQ management method (section 2.3.1) is confirmed by the case study by specifying a variety of possible data roles, such as data stewards or champions.
- Migration and integration practices as part of the criteria for DQ management in CRM D&C projects concluded from literature as listed in section 2.3.2 is confirmed by the case study as well. The practices of the *Transform* phase of the ETL process which is required for data migration and integration practices as found in literature (Thalheim & Wang, 2013; Stone et al., 2017) are confirmed by this case study. This includes the mapping of the data as stated by literature (section 2.1.2), as well as the assessment and improvement of DQ.

- Following previous item, DQ assessment as part of the criteria for DQ management in CRM D&C projects concluded from literature as listed in section 2.3.2 is included in the case study too. This includes the variety in definitions of DQ in terms of dimensions and quality targets for specific CRM D&C projects, as well as DQ measurement (section 2.2.2). Furthermore, the need for an impact evaluation for the identification of critical elements for the assessment of DQ is confirmed with terms as *top down* and *demand driven* approaches.
- DQ improvement as part of the criteria for DQ management in CRM D&C projects concluded from literature as listed in section 2.3.2 is included in the case study too. This includes the relevance of the detection of root causes of the DQ issues and ideally an impact evaluation to support a decision making process for the development of an improvement strategy (section 2.3.1, (Spruit & Pietzka, 2015)).
- Lastly, iteration as part of CRM D&C projects and DQ management as found in literature (section 2.3.1) is confirmed by the case study too, which is explained by means of the data life cycle, in which the data should be managed constantly.

In terms of new insights that emerge from the case study can be concluded that every CRM D&C project requires its own unique plan for DQ management. The case study provides insights in the uniqueness of every CRM D&C project, hence the uniqueness in necessity for DQ management per CRM D&C project. This unique DQ management plan depends on the DQ management expertise level, or maturity level, of the client as well as the client context.

Chapter 4

Design of the CRM-DQMF

In this chapter, the design of the CRM-DQMF is discussed. First, the requirements on which the design is based are discussed. Second, the PDD of the CRM-DQMF is elaborated on. The entire CRM-DQMF comprises various components, for which the design decisions are discussed separately.

4.1 Requirements Specification

Based on the triangulated results of the literature research and the case study expert interviews and documentation analysis, below requirements are identified for the CRM-DQMF. In order for the CRM-DQMF to assist CRM D&C teams effectively in the management of DQ, it should adhere to certain criteria. Those criteria ensure the management of DQ in general, as well as the management of DQ in the context of CRM D&C projects.

1 - Modularity: Due to the uniqueness of every client and project, the CRM-DQMF should be applicable in varying situations serving different needs. To serve this requirement, the CRM-DQMF exists of different components that can be separated and/or (re)combined when required. This way, the CRM D&C team can be either the executive or advising force of a component, or take a more passive role and omit the component of the CRM-DQMF, leaving the responsibility entirely to the client. By the introduction of an activity which produces a CRM D&C project specific DQ management plan as explained in the next criterion, the remaining utilisation of the CRM-DQMF is decided on. The modular visualization is facilitated by making use of situational method engineering as proposed by van de Weerd and Brinkkemper (2008).

2 - DQ Management Plan: Every project and client is unique, which means every project and client requires different DQ management practices and levels. To manage the DQ appropriately for a project, an unique DQ management plan should be established at the start of any project. It describes the roles and responsibilities of the client as well as the CRM D&C team with regards to the required DQ management practices for the CRM D&C project. The DQ management plan is established based on the scope, business goals, and budget of the specific CRM D&C project, the current DQ management maturity level of the client, as well as the DQ management goals of the client (see criterion 3 and 4).

3 - DQ Management Maturity Level: The current DQ management maturity level of the client as well as the goal DQ maturity level of the client play relevant roles for DQ management in CRM D&C projects. To determine the maturity level, a maturity

matrix is extracted from research of Spruit and Pietzka (2015), which is confirmed by and slightly edited based on findings of the documentation analysis within the case study to fit the purpose of this research. The maturity matrix can be seen in Appendix C.2. The maturity of DQ management is determined by four different capabilities: *Assessment of DQ*; *Impact on Business*; *Root causes of poor DQ*; and *DQ Improvement*. For each capability, the client can be at another maturity level, reading from lowest to highest *Initial*, *Repeatable*, *Defined Process*, *Managed & Measurable*, and *Optimized*. There does not exist a maturity level for the ignorance of the importance of DQ challenges in the maturity matrix, as this is considered as no maturity at all. On the *Initial* level, attention on the importance of DQ management has just been raised and some initial plans are developed for the investigation of the challenges concerning DQ. On the *Repeatable* level, there exist some isolated measures for the management of DQ initiated by single units or employees, without any relation amongst each other. On the *Defined Process* level, the DQ management measures taken are more aligned and their is awareness of taken measures throughout the organisation. This is the start of collaborative DQ management within the organisation for the purpose of the CRM D&C project. On the *Managed & Measurable* level the organisation has best practices for DQ management in place, following implemented frameworks and defined processes. On the final level (*Optimized)*, all required DQ management practices are optimized for the CRM D&C project beyond best practices. On this level, the benefit drawn from DQ management is at its maximum. The DQ management practices are constantly reviewed for opportunity for optimization.

4 - CRM D&C Client Context: As aforementioned, every project will require its own DQ management plan. The development of such a plan requires a definition of the business case, as this formats the required DQ management practices. The business case comprises of the clients' budget, the client's business goals with CRM D&C project, and the scope of the project. The budget indicates to which extent the client will be able to pay for assistance in DQ management services of the CRM D&C team. The business goals provide a rough indication of the need for DQ management. The scope of the project indicates which functionalities are required for reaching the business goals, and thus provides for an indication of the extent to which DQ management is required. Additionally, the client context comprises information for a reconstruction of the organisation in regards to the CRM solution, which includes business processes, the data, data policies, and data standards. Conventionally, a client context is already established by the CRM D&C team and the client as general part of the CRM D&C project.

5 - Migration/Integration: CRM D&C projects comprise the process towards a CRM solution, which typically includes the migration and/or integration of data. Therefore, the CRM-DQMF should contain guidance to integrate DQ management into migration and integration practices, including data mapping practices.

6 - Iteration: In terms of the CRM D&C project, iteration needs to take place to successfully establish migration and integration, as it was found that most CRM D&C projects follow an agile approach. Migration and integration consist of continuous gathering and refinement of business requirements, business rules, and data mappings. In terms of DQ management, this requires iterations of DQ assessment and improvement practices. Once the CRM solution is established, the CRM-DQMF

should still provide for iterations, as the assessment and improvement phases of the CRM-DQMF need to be ongoing processes, which solely end in case of the end of the data lifecycle of all concerned data within the scope of the CRM D&C solution.

7 - Business impact analysis: The business impact of poor DQ needs to be analysed, as this defines the data elements that are critical for the client's business goals and thus require DQ assessment and potentially improvement practices. Due to the variety in CRM D&C projects and clients, the eleven business impacts of poor DQ as identified by Spruit and van der Linden (2019) are recommended to be included within the impact analysis, as they are found to be applicable for a variety of industries. Furthermore, by not exclusively including monetary impact, they support an agile project approach, as agile values an emphasis on the quality, the flexibility and the customer-centricity of services (Rosing, Scheel, & Gill, 2015) and cost efficiency is not at the centre of attention in an agile project approach (Gill & Henderson-Sellers, 2006). The business impacts include *lost sales opportunities*, *customer service costs*, *customer dissatisfaction*, *lost revenue*, *operational deficiencies*, *delays in system/project deployment*, *regulatory compliance*, *poor decision making*, *lost business opportunities*, *employee moral*, and *system credibility*. For each significant business impact, a metric has to be defined to calculate its value. Values can be absolutes (e.g. monetary values), a percentage with respect to some reference tables (e.g. 3% of turnover), or a qualitative evaluation (e.g. low or high) (Batini et al., 2007).

8 - DQ Assessment: Within DQ management, DQ assessment is found to be a critical part and should take place frequently. Therefore, the CRM-DQMF should include guidance in assessment practices, containing the definition and measurement of DQ, and reporting on potential DQ issues.

9 - DQ Improvement: When data is found to be of insufficient quality, the CRM-DQMF should offer guidance in establishing an improvement strategy. For optimal DQ, an organisation needs to be aware of different reasons for poor DQ and where they are existent within the organisation, hence the root causes of DQ issues need to be analysed. The whereabouts of the weak spots should be known, as well as the reason(s) for the existence of the weak spots. This knowledge will serve as input for the establishment of an improvement strategy. To determine the improvement strategy, direct and indirect costs of DQ should be compared by making use of a cost evaluation. These costs include the costs of the business impacts of the DQ issues, as well as the costs of potential improvement practices.

4.2 CRM Data Quality Management Framework

The CRM-DQMF is composed of several phases, which can be utilized as a whole as well as on their own. A PDD of the entire complete CRM-DQMF for a typical CRM project where the migration or integration of data should be performed can be found in Figure 4.1. The phases included are *Project definition*, *Client profiling*, *Preparation*, *Migration/Integration*, *Data quality definition*, *Assessment*, and *Improvement*. The next section explains the process overview. The distinct phases along with their interdependencies are elaborated on in the following sections. The final section elaborates on the CRM-DQMF once data is successfully migrated and/or integrated, as DQ management is required to be an ongoing process in CRM and may still be part of a CRM D&C project.

4.2.1 Process Overview

As aforementioned, the entire CRM-DQMF comprises seven phases. The *Project Definition* is executed once at the beginning of every project to indicate what the project will entail in terms of DQ management. Its output is a unique DQ management plan, which determines the utilisation of the remainder of the CRM-DQMF. *Client Profiling* is performed to reconstruct the client's organisational environment with regards to the CRM solution. The output is a client profile which can be utilized for the definition of data roles, business impact analyses, and root cause detection. *Preparation* gathers the required information for the data mapping and assessment phases, which includes the definition of data roles, business requirements and business rules. *Migration/Integration* is performed to migrate and/or integrate the data with the new CRM solution. *DQ Definition* and *Assessment* are performed to define and measure the DQ, and *Improvement* is performed to improve potential DQ issues. DQ is improved by refining business requirements and business rules till DQ is determined to be of sufficient quality. Then a migration/integration plan is established and executed.

4.2.2 Project Definition

The project definition phase of the CRM-DQMF is the only phase that is executed by default for every project. As aforementioned, every client and project of a CRM D&C team is unique. Therefore, every project requires its own DQ management plan. In the project definition phase of the CRM-DQMF, the DQ management plan is defined. This definition takes several actions, which are explained in the following paragraphs.

Establish business case: A CRM business case is established to define the CRM D&C project. The activity is extracted from the blueprint as explained by the case study (section 3.3.1). This business case includes the business goals of the client and the CRM D&C project scope. The business goals and scope indicate whether data is required to be of high quality, and to what extent DQ management is of relevance. The budget of the client is determined, as this indicates the monetary boundaries of the CRM D&C project and the possible inclusion of the DQ management services of the CRM D&C team in the project proposal. Often, concessions have to be made either on DQ to deliver client experience within the constraints of costs and technique, or on the budget from the client's side.

Identify maturity level: The DQ management maturity level of the client should be taken into account for the design of a DQ management plan, as this indicates to what extent the client requires the assistance of a CRM D&C team in terms of DQ management. The activity is extracted from the insights on the DQ management expertise level of the client from the case study (section 3.3.1), and its relevance is supported by research of Spruit and Pietzka (2015) as explained in section 4.1. The DQ management maturity level of the client can be indicated by making use of the maturity matrix as can be seen in Appendix C.2, which is extracted from the research by Spruit and Pietzka (2015) and confirmed by the case study and related literature regarding best practices for DQ management methods (section 2.3.1).

Identify data quality management goal: Using the DQ management maturity matrix the client's goals of DQ management are indicated as well. As every client and

project is unique, the goals of DQ management depend on the scope of the CRM D&C project and the business goals of the client, which determines the importance of DQ management. The current maturity level of DQ management next to the goal maturity level indicates at which capabilities of DQ management the client requires to grow with regards to the CRM D&C project. This is used as input for the establishment of the DQ management plan.

Establish data quality management plan: Formatted by the business case, and using the input of the current maturity level as well as the goal maturity level of the client, the CRM D&C team develops a DQ management plan together with the client. This activity is extracted from the insights on the relevance of modularity gained from the case study in section 3.3.1. The key activity is the definition of the key roles and responsibilities for the realization of the target DQ management maturity level for each DQ management capability. It describes the DQ management services provided by the CRM D&C team, as well as the role descriptions and responsibilities of the client with regards to the management of DQ. The plan determines the remainder of the use of the CRM-DQMF. For example, the role of the D&C team regarding DQ management might end at this point, which makes the rest of the CRM-DQMF otiose for the specific project. In other cases, it can be decided that the D&C team provides advice on the steps required for DQ assessment in data migration or integration practices, after which the client takes over.

4.2.3 Client profiling

To understand and reconstruct the client's organisational environment with regards to the CRM solution, a client profile needs to be established which contains information on the data, business processes, resources, data policies, and data standards. The client profile corresponds to the blueprint as defined by the case study (section 3.3.1) and the reconstruction of the organisational environment as defined by literature (section 2.3.2). Below, the different activities that gather the required information are explained.

Identify data: The concerned data for the CRM solution need to be identified, so it is known which data should be subject to DQ management practices. The volume of the data is required in order to appropriately define a migration or integration strategy and to properly indicate the magnitude of potential DQ issues. The locations of the data are required to properly indicate where the data affects the business, and hence determine the kind of DQ issues. The type of the data is required to determine the most appropriate DQ definition and measurement techniques.

Identify concerned business processes: Business processes that are concerned with the business case as established in the project definition phase are identified. The business processes create, use, move, or modify the concerned data, and form a complicated technical and business process landscape indicating the whereabouts and purposes of the data.

Identify resources: The resources of the data need to be identified. This includes human resources, such as employees that enter the data, data sources that produce the data, and applications that utilise, move, or modify the data. The resources provide insights on the places of potential business impact caused by DQ issues.

Subsequently, it can be used as input for the development of an improvement strategy.

Identify data policies: Data policies at the client's side should be identified, as well as data policies at the side of the D&C team. It should include internal as well as regulatory data policies. They are directives that codify principles and management intent into rules that govern the data. Data policies might include, for example, rules about data classifications of criticality or General Data Protection Regulation (GDPR). The data policies are input for the definition of DQ requirements.

Identify data standard: The existing data standard for all concerned data is identified. A data standard conditions the data to ensure that it meets rules for content and format. Data standards contribute to the definition of DQ, since they provide a means for comparison. The data standard requires reviewing and refinement continuously.

4.2.4 Preparation

The preparation phase gathers the required information for the data mapping and the assessment. It is extracted from the literature on existing DQ management methods in section 2.3.1 and the findings of the case study regarding *Business Requirement* as seen in section 3.3.2, and *Data Roles* as seen in section 3.3.3.

Define usage & ownership: For all data, the data usage, data ownership, and data access need to be defined. Data usage includes all that make use of the data. Data ownership defines who is responsible for the data. Data access defines all that have access to the data.

Define business requirements: Business requirement need to be extracted from the business goals with the CRM D&C project. The business requirements describe what needs to be done to achieve the business goals.

Define data quality business rules: Designated data analysts define and refine business rules that describe expectations about the quality of the concerned data. They should be created through analysis of business processes, data policies, data standards, business impact of data, assessment reports, and common sense. Business rules are generally associated with the way data is collected or created. For example, when a client wants to send monthly newsletters to a specific sample of its customers as part of its marketing strategy in CRM, a business rule could be about the population of demographic fields such as *birth date*, or contact information fields such as *email address*. In this case, a validity rule might describe the format of the field *birth date* in "dd/mm/yy", and a completeness rule might describe the population of the field *email address* to be mandatory. The definition of business rules helps to discover pain points, what would happen in case of missing or incorrect data, or how to recognize poor DQ. The definition and refinement of business rules is an ongoing process.

4.2.5 Migration/Integration

To prepare for data migration or integration, the source system of the specific data is identified, as this determines the applicable data standards. Additionally, a mapping model is developed by defining the profile of the CRM platform and map this to the definition of the legacy CRM or the integration as stated by literature in section 2.1.2 and confirmed by the case study insights on *Data Dapping* in section 3.3.2. After data mapping, the quality of the data should be assessed, for which the responsibility lies by the defined data roles. Once the quality of the data is assessed and, when required, improved, a migration and/or integration plan is created and executed. This should be done after (some iterations of) assessment and improvement practices, since the case study indicated that, ideally, potential issues regarding DQ are resolved before migration or integration is established to prevent more significant problems.

4.2.6 Data Quality Definition & Assessment

The input for the DQ assessment phase is the gathered knowledge of the previous phases. When no DQ issues occur, migration and/or integration practices can take place. When this has already been established, the DQ definition and assessment are required to be ongoing processes, hence the process loops back to DQ definition in case of no DQ issues till the end of the data life cycle (see Figure 4.2). This iteration is extracted from literature findings on existing DQ management methods in section 2.3.1 and from case study insight in *Iteration* in section 3.3.3. When DQ issues occur, improvement practices are implemented first.

Perform impact analysis: By performing an impact analysis, critical data elements are identified. Critical data elements represent data that is of utmost importance for the achievement of the business goals. Those elements are required to comply with their DQ definitions. The result of the impact analysis is a prioritised list of data elements which can be used by the team to focus their work efforts. The relevance of an impact analysis is extracted from the insights on *Identification of critical areas* from literature in section 2.3.1 as well as the findings on *Critical Elements* in section 3.3.3 of the case study

Define data quality requirements: Quality is defined for the data, resulting in DQ dimensions and a DQ threshold. DQ dimensions enable the characterization of rules (e.g. email address must be populated) and findings (e.g. email address is 98% complete). They facilitate a mutual understanding of what is being measured. The DQ dimensions provide the basis for the definition of meaningful metrics. The DQ threshold defines the requirement belonging to the DQ dimension. This activity is extracted from literature findings on DQ definition and measurement in section 2.2.2 and case study insights on *DQ Definition* in section 3.3.3.

Define data quality metrics: Once the DQ dimensions are defined, metrics can be defined in order to quantify the findings of DQ. For example, a DQ business rule can be for the field *email address* to be mandatory, which translates into the DQ dimension *completeness*. The metric that can be used to measure the completeness of the field *email address* can be of type ratio, dividing the number of records where the field is populated by the total amount of records, and multiply this by 100 to get the

percentage of complete records. This activity is extracted from literature in section 2.2.2.

Measure data quality: DQ is measured either subjectively or objectively. The metrics are used to quantify the measurements. The output is the quantified measurements of DQ. This activity is extracted from literature findings in section 2.2.2 and case study insides on the necessity of DQ improvement, which translates into the necessity for DQ measurement.

Identify data quality issues: Based on the measurements and DQ business rules, DQ issues are identified. DQ issues are defined by setting status indicators for each DQ business rules in terms of its dimension(s) and thresholds. For example, the status indicator of the dimension of completeness for the field *email address* can be indicated *Unacceptable* when the measurement results in the threshold of below 80% complete.

Report on findings: The final output is an assessment report of the DQ and potential issues. The assessment report might offer a new perspective on the concerned data, from which new business rules could be articulated. All findings are documented in metadata. The necessity for a report on DQ is extracted from the case study findings on *Assessment* as explained in section 3.3.3.

4.2.7 Improvement

The improvement phase of the CRM-DQMF is only executed when DQ issues are reported on in the output of the assessment phase. When improvement activities have been applied, the CRM-DQMF loops back to the *Preparation* phase to review business requirements and business rules. In case of a strategy correction as part of the improvement strategy, the *Client profiling* phase should be revisited to review the organisational environment.

Perform impact analysis: The identified DQ issues are quantified and prioritized based on business impact. Business impacts include monetary costs of the poor DQ, as well as non-monetary impacts. It also takes into account the criticality of the data, the volume of the data, the number of business processes and stakeholders impacted by the issue, and the risks associated with the issue. This information is all extracted during the *Client profiling* and *Preparation* phases of the framework. The output is a ranked list of DQ issues that should be taken into account within the improvement strategy. This activity is concluded from literature findings on decision-making processes for improvement strategies (section 2.3.2) and findings on *Impact Evaluation* in section 3.3.3 from the case study.

Perform root cause analysis: Ideally, the DQ issues are remediated at their root cause. This could also mean controls and process improvements to prevent further DQ issues from happening. Therefore, a root cause analysis is performed to identify the root causes of DQ issues. This is concluded from literature findings on the relevance of the detection of root causes as explained in section 2.3.2, as well as the case study findings on *Root Cause* in section 3.3.3.

Develop improvement strategy: Based on the impact analysis, an improvement strategy is developed, evaluating the costs of the issue against the costs of the improvement actions performed by either the client or the D&C team. The improvement strategy ranks the issues that can be addressed immediately and at low costs, as well as more strategic improvements, such as root cause remediation and prevention practices. It contains improvement goals that are specific, achievable, and based on the quantification of the business impacts.

Perform improvement actions: The improvement strategy is put into practice. This might result in the revisitation of *Client profiling* or *Preparation* practices, or direct improvements in the data. In either way, assessment is performed again to assess the DQ.

4.2.8 After Migration/Integration

Once the migration or integration is established, the data life cycle does not come to an end, hence DQ is still required to be managed. In Figure 4.2 the CRM-DQMF after migration/integration is established can be seen. The phases *Project definition* and *Migration/Integration* are no longer part of the CRM-DQMF, and DQ monitoring will take place through continuous assessment (and improvement) of DQ. The refinement of business requirements and DQ business rules will be continuously as potential parts of the improvement strategy. The process solely ends in case of the data lifecycle of all data concerned with the CRM solution as explained by the case study in section 3.3.3.

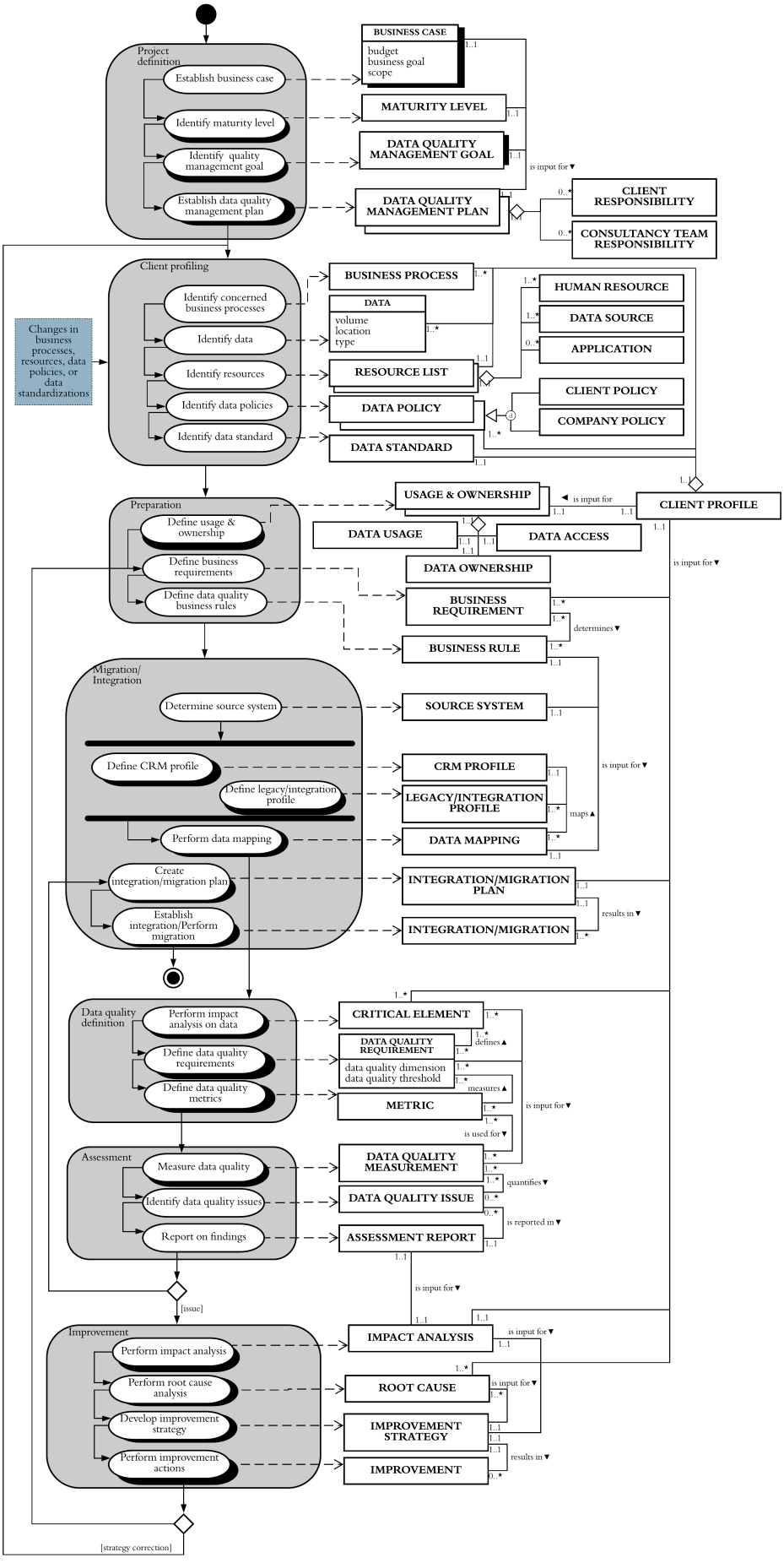

FIGURE 4.1: CRM-DQMF for a typical CRM project

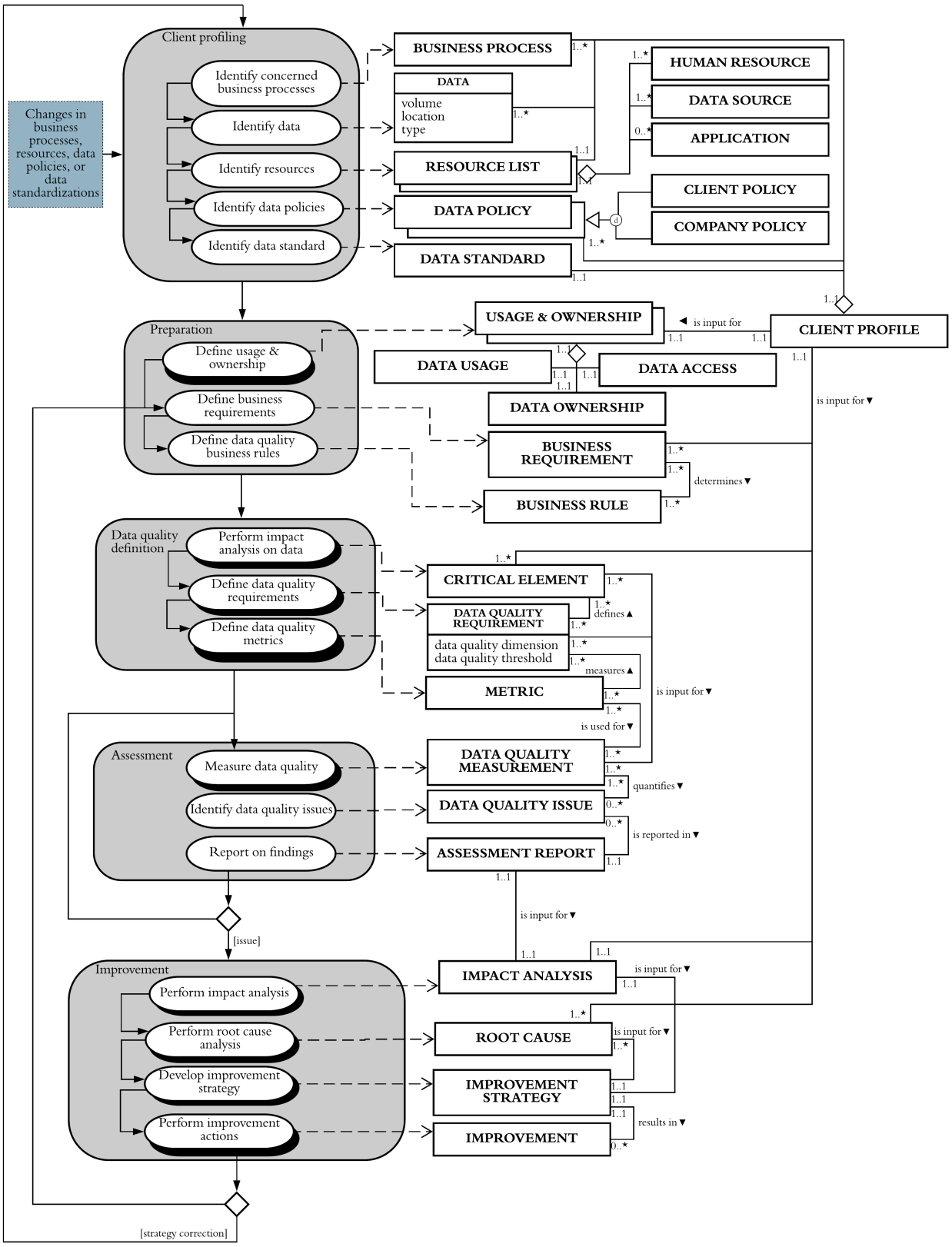

FIGURE 4.2: CRM-DQMF after migration/integration is established

Chapter 5

Validation of the CRM-DQMF

5.1 Validation method

During the validation process, no real world implementation of the CRM-DQMF is available, which means it cannot be directly investigated whether the CRM-DQMF effectively supports a D&C team in their CRM projects. To overcome this problem, a validation model of the artifact is drafted. The validation model, or design theory, consists of a description of the properties of the artifact and the interaction with the problem context (Wieringa, 2014). It facilitates a prediction of what would happen if the artifact were transferred to its intended problem context. Experts "imagine how such an artifact will interact with problem contexts imagined by them and then predict what effects they think this would have. If the predicted effects do not satisfy requirements, this is a reason to redesign the artifact." (Wieringa, 2014, p. 63). Hence, the experts form their opinion on the applicability and usefulness of the CRM-DQMF. The design theory is formed during the validation discussions, where an elaborate explanation of the CRM-DQMF is provided. Subsequently, a typical CRM D&C project is explained as the problem context. The experts are asked to think about the interaction between the CRM-DQMF and the problem context. The implementation of the CRM-DQMF into the CRM project is elaborately discussed and thus a design theory is established. By letting experts think about the implementation themselves, new insights might come up. This prevents a researcher bias.

5.1.1 Confirmatory Focus Group

The discussions that form the design theory are facilitated by the use of confirmatory focus groups. This type of focus group focuses on the utility of a designed artefact (Tremblay, Hevner, & Berndt, 2010). Focus groups facilitate group discussion(s) in which participants focus collectively upon a topic selected by the researcher. An advantage of group discussions is that is might generate ideas based on the input of others (Rabiee, 2010). Due to complex circumstances, the schedules of potential participants for the focus groups did not synchronise beneficially for the formation of focus groups with a sufficient amount of individuals. Therefore, there was chosen to perform three mini focus groups, each including two individuals (Nyumba et al., 2018). An overview of participants can be found in Table 5.1. Each participant is an employee working on CRM D&C projects.

TABLE 5.1: Participants Validation

ID	Role	Description	Experience	Session
P1	Consultant	Solution developer leading a technical team	4	1
P2	Senior Analyst (Functional)	Business related consulting in varying CRM projects	2	1
P3	Manager	Project- or delivery manager (technical), solution architect	9	2
P4	Senior Consultant (Functional)	Consulting in CRM projects	4	2
P5	Senior Consultant (Functional)	Business related consulting in CRM projects	7	3
P6	Senior Analyst	Consulting in CRM projects	2	3

5.1.2 Validation Constructs & Statements

In order to observe and measure how well the CRM-DQMF supports DQ management in CRM D&C projects, the evaluation model of Moody (2003) and research by Prat et al. (2015) are utilised. The CRM-DQMF is validated on its desired qualities (*Perceived Alignment with CRM D&C* and *Perceived Effectiveness*), its *Perceived Ease of Use*, *Perceived Usefulness*, and *Perceived Completeness*. Additionally, the CRM-DQMF is evaluated on its *Intention to Use* in order to realise the intrinsic qualities and benefits of the CRM-DQMF (Moody, 2003). The definitions of the distinct constructs are provided in the following paragraphs.

Perceived Alignment with CRM D&C Business: Perceived alignment with business is defined as the perceived congruence of the CRM-DQMF with organisations and their strategy (Prat, Comyn-Wattiau, & Akoka, 2015). In order to adapt the construct to the business context of this research, it is reformulated to *Perceived alignment with CRM D&C business*. It validates the desired qualities of the CRM-DQMF.

Perceived Effectiveness: The perceived degree to which the CRM-DQMF achieves its objectives in a real situation (Prat, Comyn-Wattiau, & Akoka, 2015; Moody, 2003).

Perceived Completeness: The perceived degree to which the structure of the artifact contains all necessary elements and relationships between elements (Prat, Comyn-Wattiau, & Akoka, 2015).

Perceived Ease of Use: The perceived degree to which experts believe that the use of the CRM-DQMF is free of effort (Moody, 2003).

Perceived Usefulness: The perceived degree to which experts believe that the CRM-DQMF is effective in achieving the intended objectives of the method (Moody, 2003).

Intention to Use: The degree to which experts have an intention to use the CRM-DQMF (Moody, 2003).

The constructs are validated by means of an interactive questionnaire containing two statements per construct. Each statement is evaluated following a 5-point Likert scale ranging from 1 (Strongly disagree) to 5 (Strongly agree). In Table 5.2, an overview of all statements can be found.

TABLE 5.2: Constructs & Statements

Construct	Statement
Alignment with CRM D&C Business	I do think that the data quality management framework is consistent with an organisation's CRM Delivery & Consultancy needs
	To my experience, the framework is not congruent/compatible with an organisation's CRM Delivery & Consultancy strategy
Perceived Effectiveness	Overall, I do think that the framework would be able to assist with in the management of data quality in CRM projects
	I do think that the framework is definitely not capable of assisting me in the management of data quality in CRM projects
Perceived Completeness	I do think that the framework is a complete overview of all data quality management practices that might occur in CRM projects
	I do miss some elements of data quality management within the framework that are applicable to CRM projects
Perceived Ease of Use	I do think that it is easy to apply the data quality management framework within CRM Delivery & Consultancy projects
	Overall, I do think that the framework is difficult to use within CRM Delivery & Consultancy projects
Perceived Usefulness	Overall, I do think that the framework is useful for CRM Delivery & Consultancy projects
	I do not believe that the framework would be useful in regards to the management of data quality in CRM projects
Intention to Use	I would intend to use the framework within CRM Delivery & Consultancy projects
	I would not use this framework for data quality management within CRM Delivery & Consultancy projects

5.1.3 Validation Session

The validation sessions are performed by making use of an online setting in Microsoft Teams. A validation session starts with an explanation of the CRM-DQMF and its objectives. Subsequently, participants are asked to think of a typical CRM

D&C project case as the problem context. The provided framework for this reads as follows:

> Company X wants to implement their marketing strategy in D365 CRM. They have an on-premise CRM system and no knowledge of the quality of the data they require for the implementation of D365 CRM Marketing. Their data is required to be of high-quality, as the consequences of poor data quality will have significant financial as well as reputational impact. As they do not have their own data quality management strategy in place, they require the expertise and consulting of Avanade.
>
> Together with the client, you walk through Project Definition phase to establish a data quality management plan, after which the data quality management framework is applied within the CRM project.

Participants are asked to think of this case in practice, as they are currently working in CRM D&C projects, which exceeds the knowledge of the researcher on contemporary CRM D&C projects in practice. The participants get the opportunity to understand the CRM-DQMF, where they can ask questions when required. Afterwards, the participants work on the case where they apply the CRM-DQMF to the CRM D&C project case collaboratively, discussing the results. This discussion is followed by an interactive questionnaire containing the statements as discussed in Section 5.1.2, which is made available through Qualtrics[1] to every participant individually. During the confirmatory focus group the statements are discussed one by one by the participants.

5.2 Validation Results

Project Definition: When discussing the *Project definition* phase of the CRM-DQMF, a question that was raised amongst the participants of validation session 3 reads as follows: "Are you showing what you want to achieve with the entire CRM project, or are you justifying to integrate data quality management?". This questions indicates that the experts are not considering the management of DQ as something that is required to be an achievement of a CRM D&C project. The validation sessions concluded that the business goal of DQ management is the improvement of DQ, while this is not part of the business goals of the client within a CRM D&C project. However, the validation sessions also confirmed findings of the case study as can be read in section 3.3.4, which state that ideally, DQ management is taken into account for every CRM D&C project in order to provide an as complete and successful delivery and consultancy for a CRM solution. Therefore, it can be concluded that for the most ideal CRM D&C project, the improvement of DQ is part of the business goals of the client, as this will result in the most optimal CRM solution. To achieve this business goal, one participant stated: "We require a process that can be used to enthuse and encourage the client to manage data quality by telling them how this can be done". The *Project definition* phase's goal is to determine to what extent DQ management services are required, while from aforementioned insights it can be concluded that the justification of the need for DQ management practices is lacking within the first design of the CRM-DQMF.

[1] https://www.qualtrics.com/support/survey-platform/survey-module/survey-module-overview/

However, in the second and third validation session, participants concluded that the steps proposed for the establishment of a unique DQ management plan at the beginning of the framework might be too expensive, as it would take a significant amount of time in relation to the magnitude of some CRM D&C projects. Only in larger projects, the establishment of a DQ management plan would be profitable. Participants provided statements such as "We don't have the money to perform those steps before the project starts to determine what efforts need to be spent on data quality management" and "For small clients it seems like a huge investment for very little data".

Client Profiling & Preparation: All validation sessions concluded that the *Client profiling* and *Preparation* phases are part of every CRM D&C project, and they agreed on the subactivities that gather the required information for successful CRM D&C as well as DQ management. To quote one participant on those phases: "This assembly is a great activity, which is definitely part of every CRM project". The concept of *Company policies* was not involved to the experience of the participants, and would rather involve *Regulator policies*, such as GDPR.

Migration/Integration: The migration and integration activities including the definition of the source system and the data mapping practices were confirmed to be required for CRM D&C projects. The participants of all validation sessions confirmed that, ideally before the creation and execution of the migration/integration plan, the data mapping and data itself are assessed and improved to eliminate DQ issues in a proactive manner. However, validation sessions 1 and 3 explained that most DQ issues and bugs are discovered after migration and/or integration takes place, as in practice it is too challenging to locate all possible DQ issues upfront. This is provided by statements such as: "Half of the bugs are discovered after migration in a dry run", and "Personally I always suggest to perform a dry run as soon as possible, to discover all bugs as quickly as possible". Participants used terms such as *dry run* and *practice round* to describe the execution of the migration/integration plan with production data within a test environment.

DQ Definition & Assessment: By each validation session the definition and assessment practices of DQ as introduced by the CRM-DQMF were perceived as important steps that are currently underexposed within CRM D&C projects. One participant stated: "Data quality definition and assessment is something we can learn from: We need to know what are the real important factors, and where do we need to put our focus. A priorities list including the *Why*". Another participant said: "Especially when clients do not have any integration yet, the establishment of a standard for data quality is the most important, and how will this be maintained. Currently, this is where mistakes are made".

Improvement: The improvement practices as introduced by the CRM-DQMF were confirmed to be applicable, and were also recognised by all participants. One participant explained: "There are various ways to ensure data quality, by automating processes, or by training your employees", which confirms the need for the development of a sufficient improvement strategy, including root cause analysis.

5.2.1 Questionnaire Results

The results of the questionnaire can be seen in Table 5.3, which are discussed individually in the following paragraphs.

The alignment with CRM D&C business and perceived effectiveness got the average scores of respectively 4,3 and 4,6 on the 5 point likert scale. Participants generally agreed that the DQ management practices are ideally integrated in CRM D&C projects, since this would provide the most complete and successful solution for the client. The DQ management practices that are proposed by the CRM-DQMF were expected to be effective in the context of CRM D&C projects, as they were interpreted as logical and would contribute to the solution to challenges that participant face during CRM D&C projects.

Perceived completeness got the score of 3,8. Participants had a hard time judging the completeness of the CRM-DQMF in terms of DQ management, as most of the participants did not explicitly consider DQ within their projects so far. However, participants judged the components of the CRM-DQMF to be logical, and could not immediately think of any components that were missing, except for a more practical approach by means of step-by-step guidance for the execution of the individual DQ management practices.

Perceived ease of use got a score of 3,8, which can be explained by the fact that most participants found the CRM-DQMF too high level, and would require more guidance for understanding how to perform the individual steps of the CRM-DQMF. To quote one participant: "I don't think it's easy to use, but I also don't think it's hard to use. It should be a way of life. which it is currently not. Lots of components are great rules for process maintenance".

Perceived usefulness was rated the score of 4,5. Participants agreed on the fact that the CRM-DQMF creates awareness on what DQ management encompasses for CRM D&C projects, from which the participants concluded the CRM-DQMF to be useful. Statements that were given on this topic read: "It is a great visual guidance to demonstrate what DQ management encompasses", and "It demonstrated what we need to take into account, but it does not provide guidance in how this should be implemented exactly, while consulting on this exact implementation should be the task of a consultant".

Intention to use scored 4,3 on the 5 point likert scale. Participants would use the CRM-DQMF as a reminder of DQ management: "I would use the framework, but rather as a tool to tell clients what they need to do themselves". The score of intention to use would increase when the CRM-DQMF provides step-by-step guidance for the individual phases and activities, as participants generally lack the knowledge or experience for performing this themselves. It would also increase when the CRM-DQMF involved some way for the CRM D&C team to inexpensively create awareness for DQ management at the client, as well as integrate the DQ management plan in the standard project proposal.

TABLE 5.3: Validation Questionnaire Results

Construct	P1	P2	P3	P4	P5	P6	Average
Alignment with CRM D&C Business	4,5	5	4,5	4,5	4	3,5	**4,3**
Perceived Effectiveness	5	5	4,5	4,5	3,5	5	**4,6**
Perceived Completeness	4,5	3	4	2,5	3	3	**3,8**
Perceived Ease of Use	4,5	3,5	4	4	2,5	3	**3,6**
Perceived Usefulness	5	4,5	4,5	4.5	3,5	5	**4,5**
Intention to Use	5	5	4	4,5	3	4,5	**4,3**

5.2.2 Main Insights

The main insights that are extracted are elaborated on in the following paragraphs.

Expensiveness: As stated by the insights on the *Project definition* phase in section 5.2, the establishment of a DQ management plan as it is presented in the first design of the CRM-DQMF might be too expensive for smaller projects. To mitigate this problem, the establishment of the DQ management plan should be more integrated in the creation of the general CRM D&C project proposal, rather than an activity on its own. However, this is not part of the CRM culture at the case study environment yet, and thus this might require some change management practices. Additionally, by performing a business impact analysis on the business goals of the client at the start of the CRM D&C project, awareness on the importance of DQ management practices at the client's side might be raised, which could result in the efforts spent on the establishment of a DQ management plan to be more profitable. When the business impacts of poor DQ appear to be negligible for the specific client and project, the remainder of the CRM-DQMF could be discarded, which also eliminates unprofitable efforts.

Awareness: The validation sessions concluded that clients might not be aware of the importance of DQ management, hence do not want to spend their budget on DQ management services of the CRM D&C team. Therefore, the impact of poor quality data on the business goals needs to be defined at the start of a project, as well as the way that high quality data will enable the business goals (DAMA, 2017). This emphasizes the importance of DQ management, creating awareness of the topic and making an indication of the need for DQ management. However, this needs to be done as inexpensive as possible in order to be profitable for the CRM D&C team, as this influences the development of the DQ management plan and thus the project proposal.

Agility: The agile project approach of CRM D&C projects did not come through cogent enough for some participants of the validation sessions. The participants argue that they would like the CRM-DQMF to guide the CRM D&C team in integrating DQ management in the concept of so-called *sprints* in agile projects, where

specific work is selected for a set period of time. First thoughts on this matter indicate that the CRM-DQMF is supposed to support an ongoing process, which could be translated into sprints, where the assessment and potential improvement of DQ iteratively takes place in every new sprint in the project.

High-Level: From each validation session can be concluded that the first design of the CRM-DQMF is too high level to put into practice as it is. Participants agreed that the CRM-DQMF creates awareness on the importance of DQ management amongst CRM D&C teams and provides relevant insights in what DQ management encompasses, rather than providing a step-by-step guidance in implementing DQ management in CRM D&C projects.

5.3 Second Iteration of the CRM-DQMF Design

From the results of the validation sessions can be concluded that the CRM-DQMF requires some modifications to be of optimal use to a CRM D&C team. The changes are implemented in the PDD based on the incremental method evolution theory as proposed by van de Weerd, Brinkkemper, and Versendaal (2007). First, the concept of incremental method evolution is explained. Second, the modifications to the CRM-DQMF are elaborated on.

5.3.1 The Incremental Method Evolution Theory

To increase the maturity of the CRM-DQMF design, the incremental method evolution approach is utilised (Weerd, Brinkkemper, & Versendaal, 2007). So-called snapshots capture the model of a process in specific moments in time, which results in the possibility to visualize the evolution of the CRM-DQMF (Weerd, Brinkkemper, & Versendaal, 2007). The differences between snapshots, called method increments, can include the following:

- Insertion: The inserted concept has a dark grey color.
- Modification: The modified concept has a dark grey color.
- Deletion: The deleted concept is visually crossed out.

5.3.2 Modification of the CRM-DQMF

In Figure 5.1 the modifications that are made to the CRM-DQMF based on the main insights of the validation sessions are visualized by means of *Modification* and *Insertion* method increments (dark grey color). Some insights that are extracted from the validation sessions are not mitigated within those changes, such as the need for change management to ensure the utilisation of the CRM-DQMF by CRM D&C teams, the need for step-by-step guidance through the individual phases, such as provided for the identification of the DQ maturity level in Appendix C.2, or the incorporation into agile project sprints. The changes that are made are elaborated on in the following.

Introduction of activity *Perform impact analysis on business goals* and concept *Business goals impact*

By defining the impact of poor quality data on the business goals, as well as the way that high quality data will enable the business goals the importance of DQ

management will be emphasized. This creates awareness on the topic for the client and makes an indication of the need for DQ management. When business impacts of poor DQ are defined to be negligible for the specific client and project, the remainder of the CRM-DQMF could be discarded. This might result in less effort by the CRM D&C team, as potential redundancy of the CRM-DQMF can be detected at an earlier point in time, making the CRM-DQMF expensive. This mitigates the validation insights of *Awareness* and *Expensiveness* as explained in section 5.2.2.

Introduction of concept *Project Proposal*

At the start of a CRM D&C project, the CRM D&C team delivers a project proposal to the client, which is the initial document that defines the project. The client profile is part of the project proposal. The DQ management plan should be integrated in the project proposal as well, where the costs for DQ management services that are deemed to be required for the DQ management goals of the client are included too. This mitigates the insight of *Expensiveness* in section 5.2.2.

Introduction of rotating icon at activity *Establish integration/Perform migration*

The validation sessions indicated that most issues concerning DQ are discovered when the migration is performed in a test environment as part of a dry run. The rotating arrow is added as an extension on the PDD notation as introduced by van de Weerd and Brinkkemper (2009). It indicates that migration practices can be performed multiple times as dry runs. After a dry run, the CRM-DQMF ends, which means the CRM-DQMF after migration/integration as can be seen in Figure 4.2 will be performed to assess the DQ when the CRM solution is "Live". When the issues as discovered by this dry run are eliminated, the migration will be established either again as a dry run, or in production. This comes forth from the validation results in *Migration/Integration*, as explained in section 5.2.

Interchange activities *Client profiling* and *Project definition*

The client profile should be established first, as this information is required for the execution of an impact analysis on the business goals as explained in the previous paragraph. For example, the volume of the concerned data and the purpose of business processes determine the impact on the business goals. As the *Client profiling* activities are explained to be part of any CRM D&C project, regardless of the involved DQ management services by the CRM D&C team, this does not make the process more expensive.

Change concept *Company Policy* into *Regulator policy*

The concept of *Company policy* is replaced by the concept of *Regulator policy*, as this is part of the involved data policies as concluded from the validation sessions. This comes forth from the validation results in *Client Profiling & Preparation*, as explained in section 5.2.

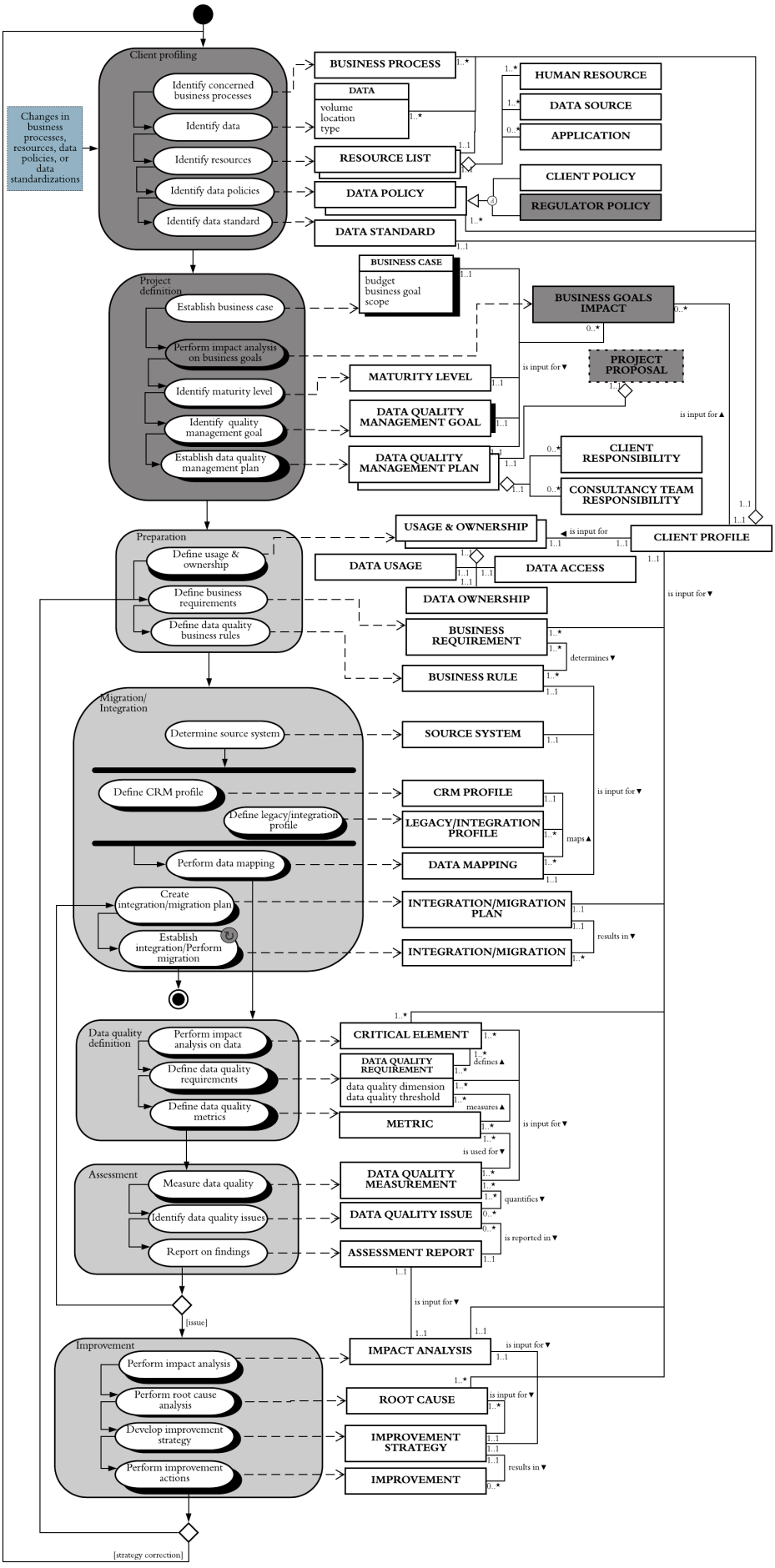

FIGURE 5.1: CRM-DQMF including changes based on validation sessions

Chapter 6

Discussion

This Chapter describes the factors that form a potential threat to the validity of this research, along with the possible diminishing of those threats. Subsequently, the limitations to this research are discussed.

6.1 Threats to Validity

The five types of validity as described by Johnson (1997) are used within this research, as they examine the validity of qualitative research. The types are: descriptive; interpretive; theoretical; internal; and external. In following paragraphs the influence of those validity types to this research is discussed.

Descriptive validity: Descriptive validity describes the factual accuracy in reporting descriptive information. For this research, only one researcher interviewed participants and conducted documentation analysis, which eliminates the possibility to achieve this validity type through investigator triangulation. To improve the descriptive validity of this research nonetheless, all conducted interviews and validation sessions were recorded to facilitate more accurate recalls of the researcher.

Interpretive validity: To achieve interpretive validity, the researcher needs to accurately describe the meaning attached by participants to the topics of the research. To ensure this type of validity, the researcher regularly incorporated participant feedback within the case study (Johnson, 1997). This is done through the utilisation of the question type *interpreting questions* as proposed by Kvale (1996) to inspect whether the interviewee's answer is interpreted correctly.

Theoretical validity: When the theoretical explanation that is developed from the research fits the data, the research achieves theoretical validity. To do so, fieldwork is incorporated in this research. This fieldwork consists of several elements, namely: participation in several training session which were facilitated to CRM D&C consultants, including D365 Fundamentals and D365 Marketing; close observation of the collaboration tool *Yammer* used by the CRM D&C community at Avanade; attending presentations on the execution of specific CRM D&C projects of several participants; joining a day of scrum meetings between Avanade and a client of the financial industry; and gain Microsoft certificates which are required/recommended for CRM D&C employees (Azure Fundamentals, D365 Fundamentals).

Internal validity: Internal validity translates into "seeking to establish a causal relationship, whereby certain conditions are believed to lead to other conditions, as distinguished from spurious relationships" (Yin, 2013, p. 46). Within this research,

method triangulation is utilised to achieve internal validity (Kaplan & Maxwell, 2005; Johnson, 1997). This means that more than one research method is used, namely literature review, expert interviews, and documentation analysis. Subsequently, data triangulation is applied by making use of multiple data sources (Johnson, 1997). Multiple expert interviews are conducted with participants form varying backgrounds. Subsequently, documentation from a variety of sources is examined for the documentation analysis.

External validity: External validity refers to the generalisability of the research findings. The case study is performed at one organisation, namely Avanade. However, various literature sources are used for the design of the CRM-DQMF, and documentation utilised for the documentation analysis originated from two additional organisations. Furthermore, the abstractness and high-level approach of the CRM-DQMF increases the external validity of the CRM-DQMF, as it facilitates generalisability to CRM D&C projects varying in client, industry, and business goals.

6.2 Limitations

Some limitations were encountered during this research, which will be discussed in the following.

First of all, due to time and resource restrictions, this research was not able to investigate the actual adoption of the CRM-DQMF within a CRM D&C project with a client of Avanade. Consequently, all conclusions are extracted from non-empirical sources, and the experience and data of experts.

Second, most participants were not consciously familiar with DQ management practices. Therefore, much effort went to exploring the expertise level of participants and being able to conduct the interviews in such a way that participants understood the concepts, while phrasing interview questions to extract required information without creating researcher bias. Additionally, the expert interviews with the participants evolved over time. The interview protocol was applied the same in every expert interview, however, insights form early interviews contributed to the ability of the researcher to phrase the questions correctly and put emphasis in the right places. Subsequently, findings from previous interviews were sometimes discussed within later interviews. There are two advantages to this: it contributes to the formulation of some question types, such as follow-up and probing questions (Kvale, 1996), and the answers to those questions weigh more when they are agreed with by other interviewees. The disadvantage is that it could create bias, as interviewees might have been pushed towards a certain direction. To mitigate this disadvantage, results from previous interviews were only provided when the interviewee already provided an answer on their own, and there existed sufficient grounds for suspecting the previous results might be applicable for the current interview as well.

Third, potential participants for the validation sessions had busy and asynchronous agendas, which made it difficult to schedule focus groups of sufficient sizes. In the end, there was chosen to perform mini focus groups consisting of two participants (Nyumba et al., 2018).

Last, the research process has been impacted by the need for all research efforts to be arranged online due to the COVID-19 regulations set by the government, the university, and Avanade. This might have resulted in less sufficient sampling results. Subsequently, it might have influenced the interpretation of the researcher, as online settings were sometimes lacking in terms of connection.

Chapter 7

Conclusions & Future Work

This Chapter describes the conclusions extracted after conducting the research. Subsequently, it describes the contribution and proposes future work that followed from conducting this research.

7.1 Research Findings

This section contains the conclusions to this research by answering the sub-questions as formulated in Chapter 1. Those conclusions lead to an answer to the main research question.

RQ1: What does the delivery and consultancy of a contemporary CRM platform consist of? The delivery and consultancy of CRM platforms involves the delivery of a CRM platform either on-premise or in the cloud, as well as consulting for the most beneficial utilisation of the CRM platform. People, processes, and technology need to be integrated, which is done by performing several smaller projects. The specific organisation needs to be examined, data needs to be migrated to the new CRM platform, and often data needs to be integrated for the optimal utilisation of the CRM platform. The ideal delivery, as well as the most beneficial utilisation of CRM D&C platforms, includes continuous management of DQ. Hence, a contemporary CRM D&C project consists of the following: an examination of the specific organisation; data migration and/or integration practices; continuous DQ management.

RQ2: How can quality of CRM data be defined? CRM platforms utilise multiple types of data: structured, semi-structured, and unstructured. Moreover, a large variety of organisations from varying industries utilises CRM platforms, which requires contemporary CRM platforms to be increasingly complex to meet all demands. To facilitate this variety in data types and demands, the definition of DQ in CRM is highly context dependent and is defined in the form the DQ dimension(s) and DQ thresholds. To acknowledge the context dependency of DQ definitions, they need to be defined for every CRM D&C project individually. To define DQ dimensions and thresholds, the business goals of the specific client and CRM D&C project need to be defined first, describing what needs to be achieved with the project. Thereafter, business requirements describe what needs to be done to achieve the business goals. Business rules describe how the business requirements can be met. DQ dimensions and DQ thresholds are aligned with business rules.

RQ3: How can CRM data quality be assessed? To include the variety of data in CRM platforms, DQ can be measured subjectively or objectively. To quantify measurements, metrics need to be defined per DQ dimension. A metric can either

be ratio, min or max operation, or weighted average. To derive an accurate picture of DQ, often different metrics have to be combined for the measurement of one DQ dimension. The DQ measurement is compared to the DQ threshold to identify DQ issues. After measurement, a DQ assessment report summarizes potential DQ issues and reports to concerned stakeholders. DQ assessment is a continuous process, which solely ends in case of the end of the data life cycle.

RQ4: What are existing CRM data quality challenges and their potential solutions? Challenges are the lack of agreement on a standard set of DQ dimensions; inconsistency in data entry and storage; data defects; condition of existing customer databases; inadequate integration of data sources; DQ deteriorating over time; high costs for maintaining DQ; and big data. Each of those challenges can be solved or mitigated by implementing an efficient DQ management strategy from the start of the CRM solution.

RQ5: What are existing data quality management approaches suitable for CRM platforms? Criteria that are extracted for the selection of DQ management methods are *upfront consideration*; *DQ assessment*; *impact analysis*; *root cause detection*; and *DQ improvement*. There exist several DQ management methods that meet the majority, or excel in some of the defined criteria for a DQ management method for CRM. They generally consist of three phases (*state reconstruction*, *assessment*, and *improvement*) in which the remaining criteria are included.

MRQ: How can a data quality management framework be designed to assist with the delivery and consultancy of CRM platforms? The designed CRM-DQMF combines scientific literature and practitioner's insights on DQ management and CRM D&C. It provides a high-level overview of DQ management practices incorporated in CRM D&C projects. With its current design, the CRM-DQMF is a tool to plan on opportunities for the incorporation of DQ management in CRM D&C projects. This incorporation involves the recognition of variety in clients and projects by the establishment of a unique DQ management plan. This plan describes to what extent DQ management services of the CRM D&C team are required for the specific project. The CRM-DQMF contains the following components:

- *Client profiling* to gather required knowledge for DQ management and the CRM solution
- *Project definition* to establish the data quality management plan for the project
- *Preparation* to define data roles and business requirements
- *Migration/Integration* to establish migration and/or integration
- *DQ definition* to define quality and metrics
- *Assessment* to measure DQ and report on findings
- *Improvement* to define an improvement strategy and improve DQ

7.2 Contribution

This research provides scientific as well as societal contributions. As shown in the problem statement, there has yet to exist a comprehensive and complete DQ management method specifically tailored for the D&C of CRM platforms. This research provides an overview of existing DQ management methods and extracts the best practices to the context of CRM D&C. It reviews what a CRM D&C project entails

and how this would benefit from DQ management. Those insights are leveraged to design the CRM-DQMF, which assists CRM D&C teams in their implementation of DQ management in CRM D&C projects. As a societal contribution, the CRM-DQMF improves DQ in CRM platforms by creating awareness of the importance of DQ management in CRM D&C and providing assistance in the application of DQ management.

7.3 Future Work

This research leaves multiple opportunities for further research. First, improvement opportunities can be found in the score of *perceived ease of use* and *perceived completeness*. Thus, further research could focus on the investigation of the usability and exhaustiveness of the CRM-DQMF. The contemporary CRM-DQMF is a high-level overview of DQ management practices incorporated in CRM D&C projects. However, as found within this study, to be of optimal use for CRM D&C teams, it requires more step-by-step guidance on how to perform or consult on the individual activities. Hence, more research can be done on the incorporation of step-by-step guidance on the execution of each activity that is included in the CRM-DQMF as well as the utilisation of the CRM-DQMF as a whole, to ensure its usability for CRM D&C teams. This might include further research to investigate on how CRM D&C teams can be directed on which components of the CRM-DQMF suit their client's situation best. Furthermore, this might include guidance on how to: establish a DQ management plan; identify maturity level and goals; perform an impact analysis; define data roles; define DQ requirements; define DQ metrics; measure and report on DQ; perform root cause analysis; and develop an improvement strategy.

Second, further empirical validation of this research is required. This research solely evaluates the design through validation sessions using a design theory and a questionnaire. Possible evaluation can include, for example, expert interviews, technical action research at actual CRM D&C projects, or surveys. Furthermore, the sampling results of the case study can be extended to experts from other fields, such as data analytics, more extensively to utilise their knowledge on the topic of DQ and include additional perspectives next to those that are utilised within this study. The results can then be used for the improvement of the CRM-DQMF.

Third, this research assumes that the participants of the case study are able to decide which practices would best fit the needs of the CRM D&C projects of the organisation. During the case study and validation sessions, some experts argued that they would require more guidance or persuasion for the application of the CRM-DQMF or DQ management practices at all. DQ management is not part of the CRM D&C culture within this case study. Hence, the adoption of the CRM-DQMF might involve change management to ensure CRM D&C teams involve DQ management practices within CRM projects more explicitly, which could result in better CRM solutions (see section 5.2.2).

Fourth, the implementation of the CRM-DQMF might be too expensive. Therefore, an opportunity for further study lies in how to integrate the establishment of a DQ management plan into the development of a CRM D&C project proposal in an as inexpensive as possible manner, as this is only implicitly mentioned within this research.

Last, this research concludes agile an customer-centric project approaches for CRM D&C projects. Hence, potential further research can review the incorporation

of the CRM-DQMF into agile project approach practices, such as sprints, as this is only implicitly mentioned within this research.

Appendix A

Data Quality Management Methods

A.1 AIMQ

AIMQ focuses on benchmarking, which is and objective and domain-independent technique for quality evaluation. The foundation is the PSP/IQ quadrant as can be seen in Table A.1. It classifies quality dimensions according to their importance from the user's and manager's perspectives into four classes, namely *Sound*, *Dependable*, *Useful*, and *Usable*. Benchmarking ranks the information within each of the classes. Gap analysis techniques are used to conduct benchmarking and interpret results. Information Quality Benchmark Gaps compares the quality values of an organisation with those of best-practice organisations. Information Quality Role Gaps compares the Information Quality (IQ) assessments provided by different organisational roles (information system professionals and the information users). A gap exists when there is a difference the professionals' and the users' assessments. When the assessment of the professionals in much higher than the assessment of the users, it indicates that the professionals are not aware of quality issues detected by users. AIMQ mainly describes assessment activities. Guidelines, techniques, and tools for improvement are not provided. The phases of AIMQ are summarized in Figure A.1.

TABLE A.1: PSP/IQ Quadrant

	Conforms to specifications	**Meets or exceeds consumer expectations**
Product Quality	Sound Information	Useful Information
Service Quality	Dependable Information	Usable Information

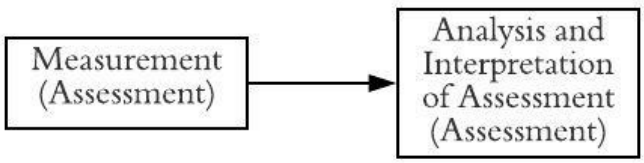

FIGURE A.1: Phases of AIMQ

Phase 1 - Measurement: The input is the defined quality dimensions according to the PSP/IQ model. The quality is measured by creating a questionnaire for measuring IQ along the important dimensions. The output is a subjective quality assessment.

Phase 2 - Analysis and Interpretation of Assessment: The input is the assessment results. The quality is measured using IQ Gap Analysis techniques and aggregation of the dimensions into the PSP/IQ quadrants. The output is improvement directives.

A.2 CDQ

Within CDQ, problem identification through interviews as well as quantitative evaluation of quality issues are performed. DQ issues are determined by means of user interviews including the identification of the relevant data and the corresponding processes. Subsequently, the DQ metrics are applied to the identified DQ dimensions.

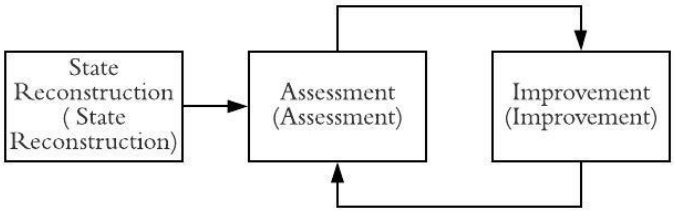

FIGURE A.2: Phases of CDQ

Phase 1 - State Reconstruction: The input is internal databases and information flows, external sources, and organisational structures and rules. State reconstructions takes place with as output a description of the organisation along its data, information flows, and processes.

Phase 2 - Assessment: The input is the organisation's description. The following steps are executed:

- Data Analysis.
- DQ Requirements Analysis.
- Measurement of quality.

The output is DQ values.

Phase 3 - Improvement: The input is the DQ values. The following steps are executed:

- Identification of root causes of errors.
- Selection of strategies and techniques.
- Evaluation of costs.

The output is the costs and benefits, the optimal improvement process, and controlled processes.

A.3 COLDQ

The goal of COLDQ is to provide a DQ scorecard supporting cost-effect evaluation of poor DQ, obtaining a quantitative assessment of the extent to which business processes are affected by poor DQ. Hence, it supports extensive cost-benefit analyses. The method provides a classification of operational, tactical, and strategic economic impacts that have to be considered. In Figure A.3 the phases are visualized, followed by a description of each phase.

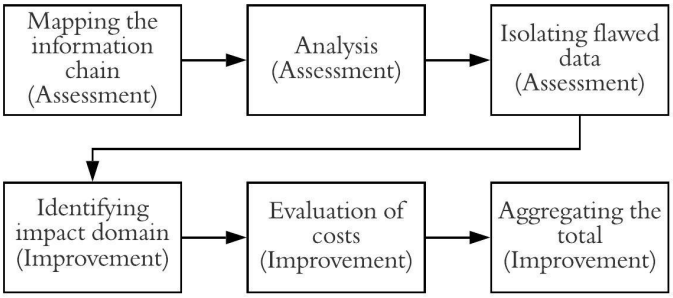

FIGURE A.3: Phases of COLDQ

Phase 1 - Mapping the information chain: The input is organisation's sources. Process modelling is executed to identify the steps through which the input data is converted into usable information. The output is an information chain and data flows.

Phase 2 - Analysis: The input is the information chain and the data flows. A DQ requirements analysis is performed by means of employee and customer interviews. The output is IQ problems.

Isolating flawed data: The input is the interviews. Process modelling is executed, with the output of identified activities attributable to poor DQ.

Phase 3 - Identifying impact domain: The input is the identified activities of previous step. Strategies and techniques are selected for the identification of impact. The output is a classification of the activities related to poor DQ.

Phase 4 - Evaluation of costs: The input is the classification. Costs are evaluated with the output of a DQ scoreboard matrix.

Phase 5 - Aggregating the total: The scoreboard matrix is the input. The costs are aggregated and the output is the poor DQ costs.

A.4 DQA

DQA is designed to guide the definition of DQ metrics. The qualification of metrics is summarized in Figure A.4. The method considers subjective as well as objective metrics.

FIGURE A.4: Phases of DQA

Phase 1 - Subjective and Objective Measurement: The input is the database(s) in use. The quality is measured by subjective as well as objective means. The outputs are the assessment results.

Phase 2 - Comparison: The input is the subjective and objective assessment results. The quality is measured by means of an analysis of comparison. The output is the discrepancies.

Phase 3 - Improvement: The input is the discrepancies. Two steps are performed within the improvement phase:

- Identification of the root causes of errors.
- Selection of strategies and techniques for improvement.

The output is the actions for DQ improvement.

A.5 DQAF

Rather than providing several steps, the assessment process according to the DQAF is based on different measurement types. It follows a cascading structure that covers the DQ aspects of data collection, processing, and dissemination. DQAF covers five quality dimensions and a set of prerequisites for DQ assessment. The dimensions recognize the characteristics of the data production systems as well as the individual data product. The cascading structure can be found in Figure A.5, with an example for the dimension *Serviceability* on the right. DQAF identifies elements that point toward DQ for each dimension. For each element, the DQAF identifies indicators. For each indicator, focal issues that are specific for the compilation of a particular dataset are addressed through key points that may be considered in identifying quality.

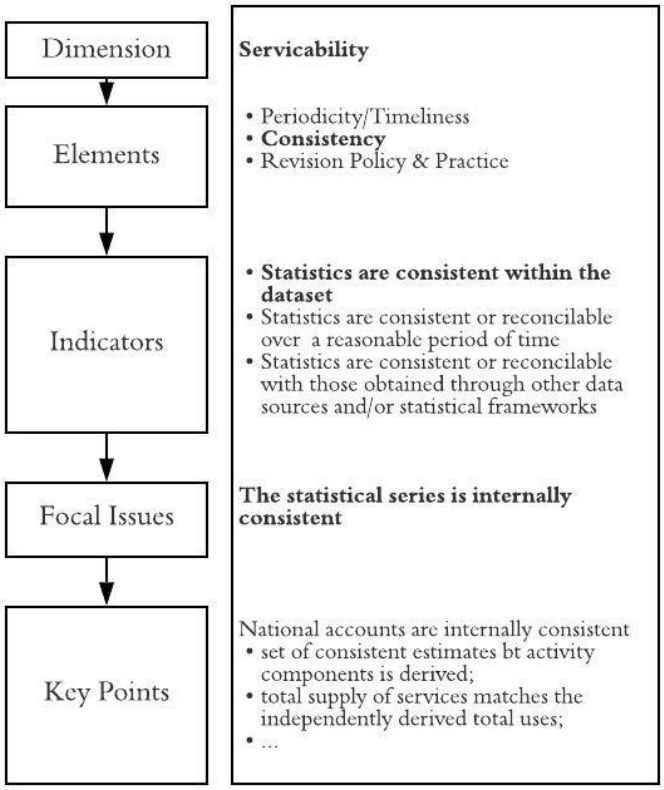

FIGURE A.5: Structure of DQAF

A.6 DQPA

DQPA is a DQ management method as an implementation of DQ assessment. According to DQPA, the ultimate point of origin of the data must be found by providing information about regarding data as an atomic values, or if it is composed data. Data origin should be part of the root cause of DQ problems. It consists of seven phases, which are visualized in Figure A.6. The distinct phases are elaborated on below the figure.

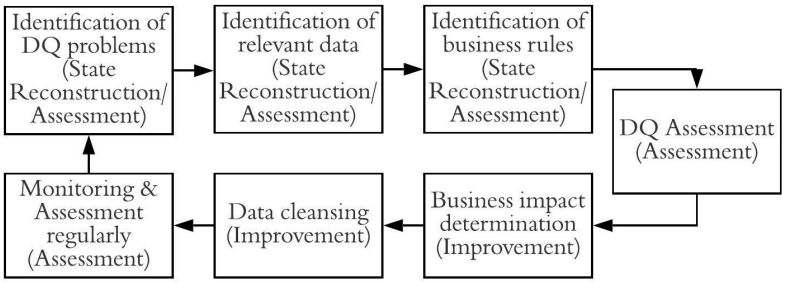

FIGURE A.6: Phases of DQPA

Phase 1 - Identification of DQ problems: DQ problems are identified by their impact on the organisation considering DQ experts, data domain experts, and end users of any level of experience. The output is the DQ problems.

Phase 2 - Identification of relevant data: The input is the DQ problems. Data that had direct impact on the organisation for an estimation of poor DQ costs is identified. The output is the relevant data.

Phase 3 - Identification of business rules: Relevant business rules are created, identified, or modified. The output is a set of relevant business rules.

Phase 4 - DQ assessment: The input is the derived data. The analysis of DQ at different level of granularity considering data source consists of several steps:

- Identification of DQ properties with as output a DQ Reference Model containing an objective and effectively set of quality criteria to provide an unbiased measure of quality to data users.
- Identification of DQ metrics with as output a Measurement Model, which assembles and extends the existing metrics for the measurement of data at different levels of granularity.
- Identification of DQ assessment methods with as output an Assessment Model, including objective and process criteria.
- Quality assessment of primary data sources by direct and indirect assessment. It corresponds to the estimation of the DQ scores of primary data sources. It is stored in Quality Metadata.
- Quality assessment of derived data using assessment by source. It required the definition and population of source metadata.
- Analysis of DQ by means of two methods. The selection of the best data sources before the query execution on the bases of its quality scores, or the comparison of DQ aggregated scores corresponding to other query plans for the same question.
- Ranking of data sources or queries using DQ scores.

Phase 5 - Business impact determination: The business impact is determined by means of a DQ comparison between the expected DQ and the actual DQ scores. The output is the feasibility of the project for cleansing and continuous assessment and the business impact in terms of operational efficiency.

Phase 6 - Data cleaning: Cleansing the data by enforcing the defined business rules, data standardization, and data matching.

Phase 7 - Monitoring & assessment regularly: The data is monitored and the assessment phase (phase 4) is repeated regularly.

A.7 HDQM

HDQM is an extension of the CDQ method towards semi-structured and unstructured relational data sources. This means that data sources that are usually neglected in DQ management methods can be considered. The method proposes to consider

all types of data in the State reconstruction phase by using a model that allows for the description of information depending on different levels of abstraction

FIGURE A.7: Phases of HDQM

Phase 1 - State reconstruction: All relevant knowledge regarding the organizational units, processes, resources, and conceptual entities are reconstructed.

Phase 2 - Assessment/Measurement: A quantitative evaluation of DQ problems is obtained. DQ dimensions are measured to assess the current level of DQ. New DQ targets are set.

Phase 3 - Improvement: Improvement activities are selected by means of DQ dimensions/costs ratio.

A.8 HIQM

HIQM is a complete method for IQ management and improvement. It is composed of eight phases, which are visualized in Figure A.8. Business processes and context are analysed, which allows for the identification of critical points in the organisation where IQ could be improved. Within the critical point, IQ blocks are inserted to monitor the information flows. A lost of suitable recovery actions and quality improvements is provided as well.

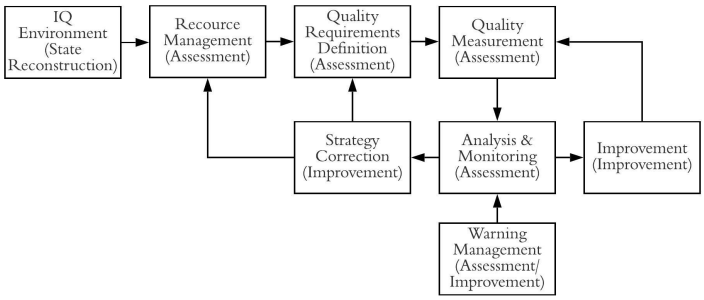

FIGURE A.8: Phases of HIQM

Phase 1 - IQ environment analysis: The business context is acquired. Data sources, processes, and stakeholders are identified, and strategic issues are considered. The output is the management plan and potential critical issues, risks critical success factors, and performance key indexes.

Phase 2 - Resource management: The input is the management plan including a definition of all involved resources. Data sources, applications and human resources that are related to each business process are identified.

Phase 3 - Quality requirements definition: It is based on the concept of the user perspective. All different information stakeholder along with their preferences and needs are considered. This includes organisation consumers, supplier consumers, and end-user consumers.

Phase 4 - Quality measurement: For each DQ dimension, a metric is identified.

Phase 5 - Analysis & Monitoring: The metrics are the input. The values associated with each dimensions are compared to the quality requirements. Causes of potential poor quality are identified along with suitable improvement actions.

Phase 6 - Data improvement: Both data-oriented and process-oriented improvement actions are considered.

Phase 7 - Strategy correction: In some situations, improvement actions imply modification at a strategic level. Changes in DQ planning systems and requirement are made in this phase.

Phase 8 - Warning Management: HIQM is the only method with a warning phase. It is an internal system for real-time data and processes monitoring.

A.9 OODA DQ

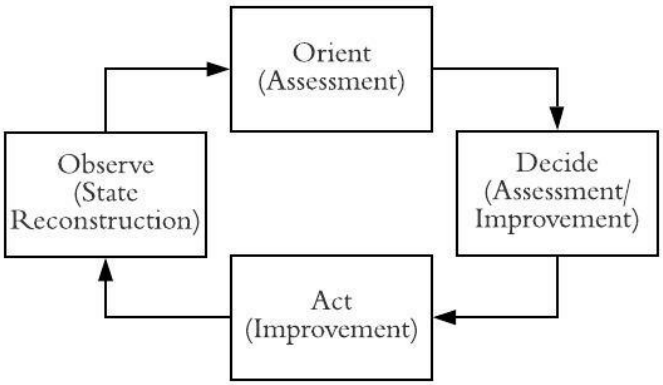

FIGURE A.9: Phases of OODA DQ

Phase 1 - Observe: Existing data is observed and data quality issues are identified by means of e.g. regular reports and dashboards.

Phase 2 - Orient: Situation analysis and judgement based on the observed results takes place. Requirements are defined.

Phase 3 - Decide: A planning policy is defined, along with strategies and tactics.

Phase 4 - Act: Corporate decisions are implemented into action plans for improvement. Monitoring takes place and the output serves as input for the observe phase.

A.10 TBDQ

TBDQ is a process-driven DQ method. Organisational processes are called process units (PU). A PU is considered as a sequence of tasks that create, read, and update data units (DU). A task is an activity that cannot be broken down into smaller activities. Each PU and each DU has a specific owner who is responsible. A team of experts in the organisation is responsible for the implementation of TBDQ.

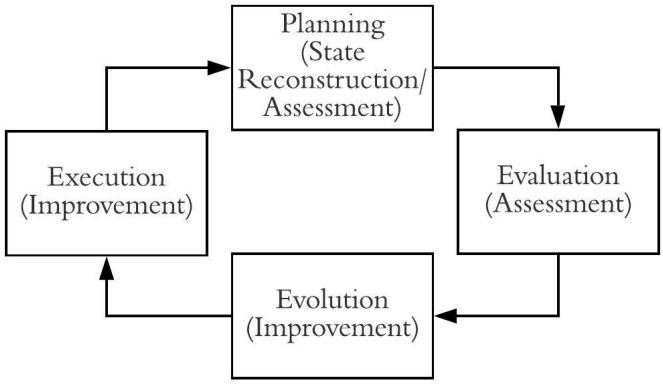

FIGURE A.10: Phases of TBDQ

Phase 1 - Planning: The objective and scope of DQ is refined. DQ dimensions, tools, risks, scheduling, and budget are identified.

Phase 2 - Evaluation: DQ dimensions are prioritized and measurement techniques are formulated. The current level of DQ is defined.

Phase 3 - Evolution: DUs are prioritized and improvement techniques are designed and selected based on costs and improvement level ratio.

Phase 4 - Execution: The improvement techniques are executed. The results are analysed.

A.11 TDQM

The goal of TDQM is to support the entire end-to-end quality improvement process, focusing on assessment and improvement. The distinct phases of TDQM can be found in Figure A.11, which illustrates a continuous quality improvement cycle. TDQM also defines the roles that are responsible for each phase: *Information suppliers* create or collect the information product; *information manufacturers* design, develop,

or maintain data and system infrastructure; *information consumers* use the data; and *information process managers* are responsible for managing the information production process. To apply TDQM, an organisation must do the following: Establish a clear understanding of its information products; establish an information product team; teach IQ assessment and management to all who is involved with the information product; and institutionalize continuous improvement.

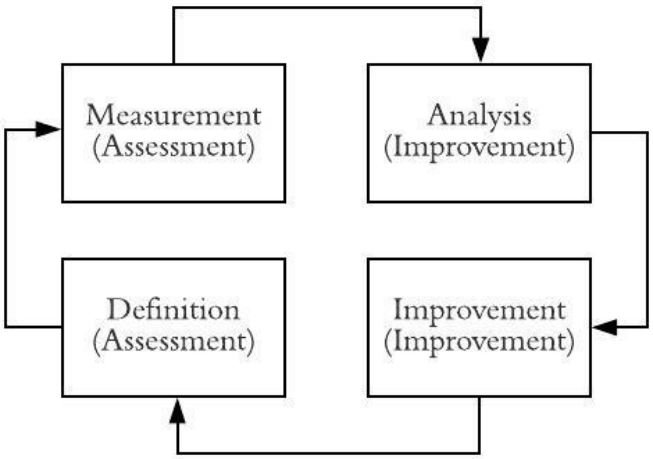

FIGURE A.11: Phases of TDQM

Phase 1 - Definition: The input is the information product. Several steps are performed:

- Data analysis for the definition of characteristics.
- DQ requirements analysis for the definition of IQ requirements.
- Process modeling for the definition of the information manufacturing system.

The output is a subjective quality assessment.

Phase 2 - Measurement: The input is identified IQ dimensions. The quality is measured by defining the IQ metrics. The output is the IQ problems.

Phase 3 - Analysis: The input are the IQ problem. The root causes of errors are identified. The output is actions for improvement.

Phase 4 - Improvement: The input for the assessment phase is the IQ metrics. The strategies and techniques for the improvement are selected, identifying key areas for improvement. The output is improvement techniques.

A.12 TIQM

TIQM assumes the consolidation of operational data sources into an integrated database, which eliminates errors and heterogeneities of source data. The focus is on management activities. The method supports cost-benefit analyses. In Figure A.12 the different phases are illustrated. Below the figure, the different phases are elaborated on.

FIGURE A.12: Phases of TIQM

Phase 1 - Assessment: The input for the assessment phase is the defined IQ criteria and requirements. Several steps are performed for the assessment:

- Data Analysis: Identify information groups and stakeholders.
- DQ requirements analysis: Assess consumer satisfaction.
- Measurement of Quality.
- Evaluation of costs.

After performing all steps, the output is an IQ assessment.

Phase 2 - Improvement: The input is the defined IQ dimensions. The quality is measured by defining the IQ metrics. The output is the IQ problems.

Phase 3 - Improvement Management & Monitoring: The input is the IQ problem. The root causes or errors are identified. The output is actions for improvement.

Appendix B

Interview Protocol

B.1 Interview and Analysis Execution

The semi-structured interviews that are conducted as part of the case study follow a protocol that is drafted from research by Halcomb and Davidson (2006), who emphasize the benefits of recordings and field notes of interviews. It is a reflexive and iterative process of data management. The analysis technique used for this research is *thematic analysis* as suggested by Clarke and Braun (2014). The resulting process of conducting and analysing the interviews and their results consists of the following steps:

Step 1 - Recording of interview and concurrent note taking: The interview is recorded and concurrent note taking consists of the researcher's impressions of an interaction rather than on transcribing verbatim sections of the interviewee's responses.

Step 2 - Reflective journalizing immediately after the interview: The field notes are reviewed and initial impressions are expanded where necessary. Additionally, the conduct of the interview is reflected on.

Step 3 - Listening to the recordings and amending/revising field notes and observations: Familiarization of the data takes place. Recordings are reviewed in consultation with the field notes to ensure that the field notes provide an accurate representation of the interview, and the researcher is aware of the breadth and depth of the data.

Step 4 - Generation of initial codes: A thorough and systematic coding of the data takes place. Codes capture features of interest of the data and can evolve during the coding process. All data relevant to each code is collated along with the generated codes.

Step 5 - Generation of candidate themes: The coded data along with the codes is examined to identify similarity and overlap. Codes are clustered together to form themes. The output is a thematic map of the analysis.

Step 6 - Review of themes: The thematic map is reviewed and required changes are made on two levels. The first level reviews the themes against the coded data to check whether collated data within the themes form a coherent pattern. The second level assesses the validity of the themes in relation to the data set and whether the themes reflect the data accurately.

Step 7 - Definition of themes: Theme definitions should tell the story of each theme, its central concept, scope and boundaries, and how it relates to the other themes and to the research question. The recordings are relistened to identify examples to demonstrate the meaning of themes from the interviewee's perspectives. A theme name should encapsulate the essence of the theme and be concise and vivid.

Step 8 - Report of the results: The results are written down in a report, including the themes, related codes, descriptions, and examples.

B.2 Informed Consent Form

See next page.


Utrecht University

Informed Consent Form

Researcher: Renee Albrecht
Organization: Utrecht University
Department: Information and Computing Sciences
Email: r.albrecht@students.uu.nl

This Informed Consent Form consists of two parts:

- Information Sheet (to share information about the research with you)
- Certificate of Consent (for signature if you agree to take part)

PART I: Information Sheet

You are asked to participate in an interview as part of a case study at Avanade on Data Quality Management within cloud-based CRM platforms. Before you decide to participate in this study, it is important that you understand why the research is being done and what it will involve. Please read the following information carefully. Please ask the researcher if there is anything that is not clear or if you need more information.

General introduction

This case study is part of a larger research that aims to create a Data Quality Management framework that assists SaaS firms with the delivery and consultancy of cloud-based CRM platforms. To this end, you will be asked questions in regard to your expertise. This interview is conducted by Renee Albrecht. The interview will be recorded, analysed and potentially used in an anonymous manner within the master's thesis of Renee Albrecht for the study Business Informatics at Utrecht University. Renee is supervised by dr. Sietse Overbeek and dr. Inge van de Weerd of Utrecht University. The interview will take approximately 60 minutes.

Burden and benefits

There will be no direct benefit to you for your participation in this study. However, the information obtained from this interview may contribute substantially to the achievement of the goals of this study. This might help with the improvement of data quality management practices for cloud-based CRM platforms.

Confidentiality

The information that we collect from this research project will be kept confidential. No-one but the involved researchers will be able to see the information that will be collected during the research. All data will be securely stored following the rules and guidelines of Utrecht University and will only be used for research purposes.

Sharing the results

The knowledge that we obtain from doing this research will be shared through scientific publications in conferences and journals. Confidential information will never be shared.

Right to refuse or withdraw

Your participation in this study is voluntary. If you decide to take part in this study, you will be asked to sign this consent form. After you sign the consent form, you are still free to withdraw without giving a reason. Withdrawing from this study will not affect the relationship you have, if any, with the researcher. If you withdraw from the study before data collection is completed, your data will be returned to you or destroyed. Note that if you withdraw after a paper regarding this project is submitted for publication, we cannot exclude your data from the current research project.

PART II: Certificate of Consent

I have read and I understand the provided information. I have had the opportunity to ask questions about it and any questions that I have asked have been answered to my satisfaction. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving a reason and without cost. I voluntarily agree to participate as a participant in this research.

Name of Participant _____

Signature of Participant _____

Date _____

B.3 Expert Interview

Introduction

My name is Renee Albrecht, and I am conducting a graduate research for the Master in Business Informatics on Utrecht University, with a case study at Avanade. The research aims to improve data quality within CRM platforms by designing a data quality management framework that assists CRM consultants with the delivery and consultancy processes of CRM platforms in order to keep the quality of data on the level required for specific organisations.

I'd like to thank you for your cooperation in this research by taking the time for this interview. You have been selected as a participant, because you have been identified as an expert in the delivery and consultancy of CRM platforms. This interview is scheduled for a duration of no longer than 60 minutes. During this time, we will discuss several open questions, which can be reviewed upfront. The questions rather function as a guidance, hence we are not bound to solely those. The aim of the questions is to gain insight into your experience and current practices regarding data quality management when carrying out your job, and gather information on the current status of data quality management practices. The results of this interview can form the input for a data quality management framework designed for CRM platform delivery and consultancy. In order to adequately process the results of this interview, I would like to audio-record the interview. Only the interviewer and supervisors on the project will be privy to the tapes. Please sign the consent form that I will deliver to you upfront. Essentially, this document states that: (1) all information will be held as confidential as possible, and (2) your participation is entirely voluntary and you may stop at any time. Thank you for agreeing to participate.

Introductory Questions

1. Who are you and how would you describe your job?
   <Understand tasks and responsibilities as a consultant>

2. How many years of experience do you have in your current job?
   <Determine experience level>

3. What project(s) did/do you work on, and in which industry? Can you shortly describe those projects?
   <Understand consultancy projects>

General Questions

4. How would you define data quality?
   <Get mutual understanding of data quality>

5. Do you consciously manage the quality of data within your projects? How?
   <Find out to what extent data quality management is already performed (proactively); Should it be managed proactively>

6. Did you ever encounter any data quality related issues within your projects? What kind of issues?

<Find challenges related to data quality; What was the solution; Are there any prevention measurements implemented afterwards; Could they have been prevented with proactive data quality management>

7. What upfront considerations do you analyse within your projects (before the SaaS CRM platform delivery)? How do different clients influence those upfront considerations?
<Define requirements for the start of a project; How does a consultant get an understanding of the business environment>

8. What types of data does the SaaS CRM platform deal with within your projects?
<Define data types>

Data Migration Questions

9. What does data migration to a SaaS CRM platform entail according to you?
<Get a mutual understanding of data migration>

10. From what types of systems do you migrate data?
<Define types of systems>

11. What steps do you take to prepare data migration to a SaaS CRM platform?
<Understand how requirements are defined for the data to be migrated; Does this differ per type of system>

12. How does the actual data migration take place?
<Discover data migration practices; When are the various practices applicable>

Data Integration Questions

13. What types of systems can be integrated with a SaaS CRM platform?
<Define types of systems>

14. What needs to be taken into consideration for integration with the SaaS CRM platform?
<Understand how requirements are defined for the data to be integrated; Does this differ per type of system; What happens when requirements for an integrated system change>

15. How does integration take place with the different types of systems?
<Discover data integration practices; When are the various practices applicable>

Data Quality Assessment Questions

16. Do you consciously take steps to assess the quality of data within the SaaS CRM platform (when it's already in production)? When?
<Find out to what extent data quality is assessed (proactively); Find out when the assessment of data quality takes place>

17. How do you decide on the meaning of quality for the different types of data to be utilized?
<Understand the data requirements gathering process; when data is of satisfactory quality>

18. How do you determine the measurement techniques and/or metrics for those requirements?
<*Understand the data measurement process*>

Data Quality Improvement Questions

19. Do you consciously take steps to improve the quality of data within your projects? When?
<*Find out to what extent data quality is improved (proactively); Find out when improvement activities are performed*>

20. How do you decide on improvement techniques when data is of unsatisfactory quality?
<*Understand when process-driven or data-driven improvement strategies are implemented; Are the costs that are related to poor data quality taken into account; Are root causes determined*>

Additional Questions

21. How do you make sure the quality data stays at a satisfactory level for the client after the project ends?
<*Understand responsibilities of a consultant; Determine whether data quality management is desired to be an active part*>

22. Do you have any additional information that might be of interest for this research?
<*Additional information that did not come up in response to the questions*>

B.4 Themes & Codes

In Table B.1, the code book for the thematic analysis can be seen. It includes descriptions of the codes, including examples extracted from the documentation analysis and the expert interviews.

TABLE B.1: Code Book

Theme	Code	Description	Example
Modularity	Client context	A description of the CRM D&C project including project scope, business goals, and budget, as well as other relevant information regarding the client's project, such as business processes and datasets	*Blueprint*; *Business Case*; P2: "together with the client we define and design business process flows"

Continued on next page

Table B.1 – continued from previous page

Theme	Code	Description	Example
	DQ management expertise level	The level of expertise and comprehension of the client on the quality of their data and DQ management	*Data Quality KPI's*; P3: "In the ideal case, organisations already have an authority in place that takes care of data quality matters. However, this varies per organisation and industry. The interference of Avanade depends on the arrangements with the client"
CRM delivery & consultancy	Business requirements	Business requirements describe what needs to be done to achieve the business goal(s)	Send an email to a certain sample of end-customers once a month
	Data mapping	The mapping of two data models to ensure correct migration and/or integration of data	P4: "It is important to understand what the data means, how this exists in the legacy system, and how this can be used in the new system"
	Agility	The ability to respond faster to changes coming from customers and market through continuous delivery in smaller iterations, and deliver the CRM solution according to the needs and preferences of the customer (Tseng & Lin, 2011; Liang & Tanniru, 2006)	P3: "We act on behalf of the client"; Continuous definition and refinement of business requirements, business rules, and data mapping of the CRM solution
Data quality management	Iteration	The frequent execution of assessment and improvement practices to ensure the quality of data	*Monitoring*; P12: "I think data quality should be measured frequently in any case"
	DQ Definition	The definition of DQ dimensions as well as DQ thresholds	P3: "Data quality includes completeness, accuracy, and relevancy. The quality depends on the system and the dataset"

Continued on next page

Table B.1 – continued from previous page

Theme	Code	Description	Example
	Critical element	A data element that is of utmost importance to the success of the CRM D&C project solution and has to comply with defined DQ requirements	*Top Down* or *Demand Driven* approach
	Data roles	The roles that describe the responsible people of the data	*Chief Digital/Data Officer*; *Data Steward*; *Champion*; P6 "We always try to assign someone who is responsible for the data at the client's side"
	Impact evaluation	Analysis of the business impacts of poor DQ	P14: "An impact analysis should be included in the decision making of what data quality means"
	Assessment	The measurement of and reporting on the state of DQ	*Transform*; *Reporting*
	Root cause	The detection of the root causes of DQ issues to resolve them at the source	*Detection of improvement location*; *Prevention Control*; P9: "The causes of data quality issues are crucial"
	Improvement	The definition and execution of improvement strategies, either data-driven or process-driven	*Resolution*; *Fuzzy Matching*; *Removing Nulls*; *Process Improvement*

Appendix C

CRM-DQMF Design

C.1 CRM-DQMF Activity & Concept Table

This section introduces the activity and concept table belonging to the PDD as shown in 4. First, the activity table is provided. Second, the concept table can be seen.

C.1.1 Activity Table

The activity table in Table C.1 provides an overview of the activities that are part of the CRM-DQMF.

TABLE C.1: Activity Table of CRM-DQMF

Activity	Sub-activity	Description
Project Definition	Establish Business Case	The CRM project is defined by establishing a BUSINESS CASE containing business goals, scope, and budget
	Identify maturity level	The MATURITY LEVEL of the client in DQ management is identified using the maturity matrix
	Identify data quality management goals	The client's DQ MANAGEMENT GOALs are identified using the maturity matrix
	Establish data quality management plan	A DATA QUALITY MANAGEMENT plan is developed based on the input of the BUSINESS CASE, the MATURITY LEVEL, and the DQ MANAGEMENT GOALs.
Client profiling	Identify concerned business processes	BUSINESS PROCESSes that create, use, move, or modify the DATA concerned with the proposed CRM solution are identified
	Identify data	The DATA that is concerned with the CRM solution, hence will be subject to DQ management
	Identify resources	The HUMAN RESOURCEs, DATA SOURCEs, and APPLICATIONs that create, use, move, or modify the DATA are identified to use as input for impact analysis and the development of an IMPROVEMENT STRATEGY

Continued on next page

Table C.1 – continued from previous page

Activity	Sub-activity	Description
	Identify data policies	The DATA POLICY is identified, which is used as input for the definition of DQ
	Define data standard	A DATA STANDARD for all concerned DATA is defined to ensure the DATA meets an agreement on rules for content and format
Preparation	Define business requirements	Define BUSINESS REQUIREMENTs to describe how to reach the BUSINESS GOALs
	Define data quality business rules	Rules are defined that describe what is expected of the DQ in order to meet the BUSINESS REQUIREMENTs
	Define usage & ownership	Determine the usage, availability, and responsibles for the DATA
Migration/ Integration	Determine source system	Identify the SOURCE SYSTEM of the DATA to use as input for the DATA MAPPING and the definition of DQ
	Define CRM profile	The profile of the CRM solution is defined to map against the profile of the legacy system or the system to be integrated
	Define legacy/integration profile	The profile of the legacy system or system to be integrated is defined to map against the profile of the CRM solution or the system to be integrated
	Perform data mapping	The DATA of the two profiles is mapped as a preparation for INTEGRATION/MIGRATION
	Create migration/integration plan	A strategy for migrating the DATA or establishing an INTEGRATION is developed, taking into account the volume of the DATA
	Establish migration/integration	The MIGRATION is performed and the INTEGRATION is established, resulting in the CRM solution
Data quality definition	Perform impact analysis	An impact analysis is performed to identify CRITICAL ELEMENTs of the CRM solution that require DQ management, taking into account various business impacts
	Define data quality requirements	DQ is defined by DQ dimensions and DQ thresholds
	Define data quality metrics	The METRICs that are used to quantify the DATA QUALITY MEASUREMENTs of the DQ dimensions are defined
Assessment	Measure data quality	The DQ is measured subjectively or objectively, using the METRICs resulting in quantitative DATA QUALITY MEASUREMENTs
	Identify data quality issues	The DQ MEASUREMENTs are benchmarked against the DQ threshold to identify DATA QUALITY ISSUEs

Continued on next page

Table C.1 – continued from previous page

Activity	Sub-activity	Description
	Report on findings	The DATA QUALITY MEASUREMENTs, including the DATA QUALITY ISSUEs are reported on to concerned stakeholders
Improvement	Perform impact analysis	The business impacts of the DATA QUALITY ISSUEs is measured to create a PRIORITIES LIST of the DATA QUALITY ISSUEs that require a solution
	Perform root cause analysis	A ROOT CAUSE ANALYSIS is performed to find the root causes to the DATA QUALITY ISSUEs
	Develop improvement strategy	An IMPROVEMENT STRATEGY is developed taking taking into account the business impacts of poor DQ, as well as the costs of potential IMPROVEMENT actions
	Perform improvement actions	The IMPROVEMENT actions as defined by the IMPROVEMENT STRATEGY are performed, resulting in either process-driven or data-driven IMPROVEMENTs to improve DQ

C.1.2 Concept Table

The concept table of the CRM-DQMF in Table C.2 provides an overview of the deliverables that are produced by the activities.

TABLE C.2: Concept Table of CRM-DQMF

Concept	Description
BUSINESS CASE	The BUSINESS CASE contains relevant information for the definition of the specific CRM D&C project, including the client's budget, the business goals, and the scope of the project
MATURITY LEVEL	An indication of the client's level of expertise in DQ management
DATA QUALITY MANAGEMENT GOAL	The client's goals with regards to DQ management
DATA QUALITY MANAGEMENT PLAN	The plan that contains the extent of integration of DQ management in the CRM D&C project
CLIENT RESPONSIBILITY	The responsibilities of the client with regards to DQ management practices included in the DATA QUALITY MANAGEMENT PLAN
CONSULTANCY TEAM RESPONSIBILITY	The responsibilities of the D&C team with regards to DQ management practices included in the DATA QUALITY MANAGEMENT PLAN
BUSINESS PROCESS	The BUSINESS PROCESSes that are concerned with the proposed CRM solution

Continued on next page

Table C.2 – continued from previous page

Activity	Description
DATA	The DATA that is concerned with the proposed CRM solution, including its volume, type, and location
RESOURCE LIST	The RESOURCEs of the DATA that is concerned with the proposed CRM solution
HUMAN RESOURCE	Human actors that create or utilize the concerned DATA
DATA SOURCE	The place where the DATA is created
APPLICATION	APPLICATIONs that use, move, or edit the DATA
DATA POLICY	Existing procedures for the management of the DATA
CLIENT POLICY	DATA POLICY on the client's side
COMPANY POLICY	DATA POLICY on the CRM D&C team's side
DATA STANDARD	Agreement on the representation, format, and use of the DATA
BUSINESS REQUIREMENT	The BUSINESS REQUIREMENTs that are required to reach the BUSINESS GOALs
BUSINESS RULE	The rules that describe what the DATA should look like to meet the BUSSINESS REQUIREMENTs
USAGE & OWNERSHIP	Information on the USAGE & OWNERSHIP of the DATA
DATA USAGE	Information on by who, by what, and where the DATA should be used
DATA OWNERSHIP	Information on the responsible people and systems of the DATA
DATA ACCESS	Information on for who, for what, and when the DATA should be available
SOURCE SYSTEM	The source of the DATA is used as input for DATA MAPPING, as this determines the DATA STANDARD
CRM PROFILE	The profile of the new CRM solution, which is mapped against the profile of the legacy system or system to be integrated
LEGACY/ INTEGRATION PROFILE	The profile of the legacy system or the system to be integrated, which is mapped against the profile of the CRM solution
DATA MAPPING	The mappings between the DATA elements of the two different systems (CRM solution and legacy system or integration
INTEGRATION/ MIGRATION PLAN	Contains the guidelines for succesfull migration and/or integration establishment for the CRM, which is influenced by the volume of the DATA
INTEGRATION/ MIGRATION	The actual MIGRATION/INTEGRATION of the DATA
CRITICAL ELEMENT	An impact analysis is performed to detect CRITICAL ELEMENTs, which is DATA that is of utmost importance for the achievement of the BUSINESS GOALs and thus requires assessment

Continued on next page

Table C.2 – continued from previous page

Activity	Description
DATA QUALITY REQUIREMENT	The definition of DQ as a characterizations of BUSINESS RULEs in the form of DQ dimensions and DQ thresholds, providing a basis for METRICs
METRIC	METRICs are defined based on DATA QUALITY DIMENSIONS to measure DQ quantitatively
DATA QUALITY MEASUREMENT	The quantitative DATA QUALITY MEASUREMENTs by METRICs
DATA QUALITY ISSUE	The DATA QUALITY ISSUEs that are concluded from the DATA QUALITY MEASUREMENTs
ASSESSMENT REPORT	A report on the quality of the DATA, including the DATA QUALITY ISSUEs that require attention from concerned stakeholders
IMPACT ANALYSIS	An IMPACT ANALYSIS is performed to determine the business impacts of the reported DATA QUALITY ISSUEs to be able to make a prioritisation on the DATA QUALITY ISSUEs that require attention and to serve as input for the development of an IMPROVEMENT STRATEGY
ROOT CAUSE ANALYSIS	The ROOT CAUSEs of the DATA QUALITY ISSUEs are investigated to serve as input for the development of an IMPROVEMENT STRATEGY
IMPROVEMENT STRATEGY	The measures that will taken as a consequence of the DATA QUALITY ISSUEs in the ASSESSMENT REPORT, taking into account the costs of improvement as well as the costs of poor DQ
IMPROVEMENT	The required alterations that are made based on the IMPROVEMENT STRATEGY, either process-driven or data-driven

C.2 CRM-DQMF Support

Below document serves as a guide for the execution of the activities for the definition of the current DQ management maturity level as well as the DQ management maturity level goals of the client. It elaborates on the key activities and provides a maturity matrix that can be used as a tool within those activities.

Maturity Level & Goals

Guide for determination of the data quality management maturity level and goals

This guide supports the second and third activity of the *Project Definition* phase of the CRM-DQMF.

The key activities for the identification of the current maturity level are:

1. Identify and validate current-state Data Quality Management capabilities and maturity levels through interviews, document study, and workshops.
2. Benchmark with the maturity level model as shown below.
3. List requirements and business objectives and pinpoint issues and challenges in realizing data quality management objectives.

The key deliverable is a maturity assessment.

Input required from the client:

1. Current data quality management capabilities.
2. Current data quality management initiatives.
3. Existing data quality (management) issues.
4. Work done so far.

Input required from D&C team:
Data quality management best practices using a data quality management maturity template.

The maturity matrix as can be seen on the next page serves as a data quality management maturity template. The maturity level is defined per data quality management capability.

The key activities for the identification of the client's data quality management goal are:

1. Identify potential value of data quality management for CRM data for the client.
2. Agree upon the required data quality management objectives.
3. Prioritize the capabilities for action.

The key deliverable is a target data quality management maturity level, for which the maturity matrix on the next page can serve as well.

Below figure provides an example of how the maturity matrix can be used to visualize the current maturity level along the target maturity level.

	Initial	Repeatable	Defined process	Managed & Measurable	Optimized
Assessment of data quality	There is a feeling about data being of good or bad quality	It is clearly stated which aspects are part of data quality and need to be measured in terms of assessing data quality	Data quality is defined regarding the requirements of different stakeholders	Data quality is measured objectively and for each piece of data it is known which quality it has	The data quality assessment is conducted frequently for all relevant data
Impact on business	The client knows that quality issues in certain data will impact the business from a reputational point of view	The client knows that quality issues in certain data will impact specific parts of the business as direct monetary loss	The client knows how poor CRM data quality impacts the business from a monetary perspective	The client knows how poor CRM data quality impacts the business from a non-monetary perspective	The client can state how insufficient data influences the business in monetary and non-monetary terms and can classify this in financial arguments
Root causes of poor data quality	The client is aware of the fact that there are different reasons for poor data quality	The client can state which reasons for poor data quality occur in the organisation	There are patterns investigated about poor data quality	The employees are aware of the root causes of poor data quality in their daily work and the consequences thereof	The client is aware of different reasons for poor data quality and where they are existent in house.
Improvement	The client figures out areas in which the data quality is not sufficient	There is awareness of the importance of high quality data in terms of efficiency and effectiveness	The client has a benchmarking system in place to assess whether the data quality is sufficient or not	Improvement measures are installed to improve The quality of data	The client regularly assesses the data quality along the Benchmarking system and ensures that the data quality stays within the defined quality

Maturity Matrix extracted from Documentation Analysis at Avanade and Spruit, M. and K. Pietzka (2015). "MD3M: The master data management maturitymodel". In: Computers in Human Behavior51, pp. 1068–1076.DOI:http://dx.doi.org/10.1016/j.chb.2014.09.030.

Bibliography

Ahmad, N., Q.N. Naveed, and N. Hoda (2018). "Strategy and procedures for Migration to the Cloud Computing". In: *International Conference on Engineering Technologies & Applied Sciences* 5.

Ali, H., R. Moawad, and A.A.F. Hosni (2017). "A Cloud Interoperability Broker (CIB) for data migration in SaaS". In: *Future Computing and Informatics Journal*. DOI: 10.1016/j.fcij.2017.03.001.

Balachandran, B.M. and S. Prasad (2017). "Challenges and Benefits of Deploying Big Data Analytics in the Cloud for Business Intelligence". In: *International Conference on Knowledge Based and Intelligent Information and Engineering Systems* 12, pp. 1112–1122.

Ballou, D.P. et al. (1998). "Modeling information manufacturing systems to determine information product quality". In: *Management Science* 44.4, pp. 462–484.

Batini, C. et al. (2006). "A comprehensive data quality methodology for Web and structured data". In: *Proceedings of the first International Conference on Digital Information Management*, pp. 448–456.

Batini, C. et al. (2007). "A Framework And A Methodology For Data Quality Assessment And Monitoring". In: *Proceedings of the 12th International Conference on Information Quality*.

Batini, C. et al. (2009). "Methodologies for Data Quality Assessment and Improvement". In: *ACM Comput. Surv.* 41.3. DOI: 10.1145/1541880.1541883.

Batini, C. et al. (2011). "A data quality methodology for heterogeneous data". In: *International Journal of Database Management Systems* 3.11, pp. 60–79.

Batini, C. et al. (2015). "Data quality to big data quality". In: *Database Management* 26.1, pp. 60–82.

Baumeister, R.F. (2013). *Writing a Literature Review.* 2nd ed. Springer and Business Media. Chap. 8, pp. 119–132. DOI: 10.1007/978-1-4614-3994-3.

Baumeister, R.F. and M.R. Leary (1997). "Writing narrative literature reviews." In: *Review of General Psychology* 1.3, pp. 311–320. DOI: 10.1037/1089-2680.1.3.311.

Benbasat, I., D.K. Goldstein, and M. Mead (1987). "The case research strategy in studies of information systems". In: *MIS Quarterly* 11.3, pp. 369–386. DOI: 10.2307/248684.

Birthare, S.K. and R.N. Sharma (2020). "Study on migration of on-premise ERP to SaaS product". In: *International Research Journal of Modernization in Engineering Technology and Science* 2.12.

Bowen, G. (2009). "Document analysis as a qualitative research method". In: *Qualitative Research Journal* 9, pp. 27–40. DOI: 10.3316/QRJ0902027.

Brinkkemper, S. (1996). "Method engineering: engineering of information systems development methods and tools". In: *Information and software technology* 38.4, pp. 275–280.

Bryman, A. (2016). "Social reseaqrch methods". In:

Budgen, D. and P Brereton (2006). "Performing systematic literature reviews in software engineering". In: *Icse '06 proceedings of the 28th international conference on software engineering*. DOI: 10.1145/1134285.1134500.

Buttle, F. and S. Maklan (2019). *Customer Relationship Management: Concepts and Technologies*. 4th ed. Routledge. ISBN: 9781351016551.

Cappiello, D., P. Ficiaro, and B. Pernici (2006). "HIQM: A methodology for information quality monitoring, measurement, and improvement". In: *Advances in Conceptual ModelingTheory and Practice*. Springer, pp. 339–351.

Cichy, C. and S. Rass (2019). "An Overview of Data Quality Frameworks". In: 7, pp. 2169–3536. DOI: 10.1109/ACCESS.2019.2899751.

Clarke, V. and V. Braun (2014). "Thematic Analysis". In: *Teo T. (eds) Encyclopedia of Critical Psychology*, pp. 1947–1952. DOI: 10.1007/978-1-4614-5583-7_311.

Creswell, J.W. and C.N. Poth (2018). *Qualitative inquiry and research design: Choosing among five approaches*. 4th ed. Sage.

Cricelli, L. et al. (2019). "Searching for the one: Customer relationship management software selection". In: *Journal of Multi-Criteria Decision Analysis*, pp. 173–188. DOI: 10.1002/mcda.1687.

Cruz-Jesus, F., A. Pinheiro, and O. Tiago (2019). "Understanding CRM adoption stages: empirical analysis building on the TOE framework". In: *Computers in Industry* 109, pp. 1–13.

DAMA, International (2017). *DAMA-DBoK: Data Management Body of Knowledge*. Vol. 2. Technics Publications. ISBN: 9781634622363.

Davenport, T. and J Harris (2017). *Competing on Analytics: Updated, with a New Introduction: The New Science of Winning*. Harvard Business Review Press. ISBN: 9781633693739.

Eckerson, W.W. (2002). "Data warehousing special report: Data quality and the bottom line". In: *Achieving business success through a commitment to high quality data*.

English, L. (1999). "Improving Data Warehouse and Business Information Quality". In:

Erceg, V., L. Šereš, and T. Zoranović (2019). "Data Quality Assessment". In: *International conference on Applied Internet and Information Technologies*, pp. 112–116.

Even, A., G. Shankaranarayanan, and P.D. Berger (2010). "Evaluating a model for cost-effective data quality management in a real-world CRM setting". In: *Elsevier* 50, pp. 152–163.

Foss, B. et al. (2002). "Managing the quality and completeness of customer data". In: *Journal of Database Marketing* 10 (2), pp. 139–158.

Francisco, M.M.C. et al. (2017). "Total Data Quality Management and Total Information Quality Management Applied to Costumer Relationship Management". In: *ICIME 2017: Proceedings of the 9th International Conference on Information Management and Engineering*, pp. 40–45. DOI: 10.1145/3149572.3149575.

Garousi, V., M. Felderer, and M.V. Mäntylä (2016). "The need for multivocal literature reviews in software engineering: Complementing systematic literature reviews with grey literature". In: *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering*. DOI: 10.1145/2915970.2916008.

Gartner (2018). *Poor quality data weakens an organization's competitive standing and undermines critical business objectives*.

Gill, A. and B. Henderson-Sellers (2006). "Measuring agility and adaptibility of agile methods: a 4 dimensional analytical tool". In: *The IADIS international conference on applied computing*.

Goldenberg, J. (2008). "CRM in Real Time: Empowering Customer Relationships". In: 18.

Hai, H and S. Sakoda (2009). "SaaS and Integration Best Practices". In: *Fujitsu Science Technology Journal* 45.3.

Halcomb, E.J. and P.M. Davidson (2006). "Is verbatim transcription of interview data always necessary?" In: *Applied Nursing Research* 19, pp. 38–42. DOI: 10.1016/j.apnr.2005.06.001.

Hanson, G. (2020). "State of CRM Data Management 2020 Report". In:

Heinrich, B. et al. (2018). "Requirements for Data Quality Metrics". In: *J. Data and Information Quality* 9.2, pp. 45–61. DOI: 10.1145/3148238.

IBM (2018). *Extracting business value from the 4 V's of big data*.

IDC (2020). "Global DataSphere". In: *IDC*.

Johnson, R.B. (1997). "Examining the validity structure of qualitative research". In: *Education* 141 (2).

Kaplan, B. and J.A. Maxwell (2005). *Qualitative Research Methods for Evaluating Computer Information Systems*. Springer. DOI: 10.1007/0-387-30329-4_2.

Kitchenham, B. (2004). "Procedures for performing systematic reviews". In: *Keele, UK, Keele University* 33.TR/SE-0401.

Klein, H.H.K. and M. Myers (1999). "A set of principles for conducting and evaluating interpretive field studies in information systems". In: *MIS Quarterly* 13.1, pp. 67–93. DOI: 10.2307/249410.

Kvale, S. (1996). "Interviews: An introduction to qualitative research interviewing." In: *Thousand Oaks*.

Laney, D. B. (2018). *Infonomics: How to monetize, manage, and measure information as an asset for competitive advantage*. Taylor & Francis Group. ISBN: 9781315108650.

Lee, Y.W. et al. (2002). "AIMQ: A methodology for information quality assessment". In: *Information Management Horizons* 40.2, pp. 133–146.

Levy, Y. and T.J. Ellis (2006). "A systems approach to conduct an effective literature review in support of information systems research". In: *Information Science Journal* 9, pp. 181–212. DOI: 10.1049/cp.2009.0961.

Liang, Ting-Peng and Mohan Tanniru (2006). "Special Section: Customer-Centric Information Systems". In: *Journal of Management Information Systems* 23.3, pp. 9–15. DOI: 10.2753/MIS0742-1222230301.

Loshin, D. (2004). "Enterprise Knowledge Management - The Data Quality Approach". In: *Series in Data Management Systems*. Morgan Kaufmann, pp. 448–456.

Marsh, R. (2005). "Drowning in dirty data? It's time to sink or swim: A four-stage methodology for total data quality management". In: *Journal of Database Marketing Customer Strategy Management* 12.2, pp. 105–112.

Missi, F., S. Alshawi, and G. Fitzgerald (2005). "Why CRM Efforts Fail? A Study of the Impact of Data Quality and Data Integration". In: *Hawaii International Conference on System Sciences* 38. DOI: 10.1109/HICSS.2005.695.

Moody, D.L. (2003). "The method evaluation model: a theoretical model for validating information systems design methods". In: *ECIS 2003 proceedings* 79.

Nagle, T., T. Redman, and D. Sammon (2020). "Assessing data quality: A managerial call to action". In: *Business Horizons* 63, pp. 325–337. DOI: https://doi.org/10.1016/j.bushor.2020.01.006.

Nam, D., J. Lee, and H. Lee (2019). "Business analytics use in CRM: A nomological net from IT competence to CRM performance". In: *International Journal of Information Management* 45, pp. 233–245.

Natchiar, S.U. and S. Baulkani (2014). "Customer Relationship Management Classification Using Data Mining Techniques". In: *International Conference on Science Engineering and Management Research (ICSEMR)*. DOI: 10.1109/ICSEMR.2014.7043662.

Negahban, A., D.J. Kim, and C. Kim (2016). "Unleashing the Power of mCRM: Investigating Antecedents of Mobile CRM Values from Managers' Viewpoint". In:

International Journal of Human-Computer Interaction 32.10, pp. 747–764. DOI: 10.1080/10447318.2016.1189653.

Nguyen, B. and D. S. Mutum (2012). "A review of customer relationship management: Successes, advances, pitfalls and futures". In: *Business Process Management Journal* 18, pp. 400–419.

Nyumba, O. et al. (2018). "The use of focus group discussion methodology: Insights from two decades of application in conservation". In: *Methods in Ecology and Evolution* 9 (1). DOI: 10.1111/2041-210X.12860.

Otto, B. et al. (2007). "Towards a Framework for Corporate Data Quality Management". In: *Australasian Conference on Information Systems*.

Petrović, M. (2020). "Data Quality in Customer Relationship Management (CRM) - Literature Review". In: *Strategic Management* 25.2, pp. 40–47. DOI: 10.5937/StraMan2002040P.

Pilar Angeles, M. del and F. García-Ugalde (2009). "A data quality practical approach". In: *International Journal of Advanced Software* 2.2, pp. 259–274.

Pipino, L.L., Y.W. Lee, and R.Y. Wang (2002). "Data Quality Assessment". In: *Communications of the ACM* 45.4. DOI: 10.1145/505248.506010.

Prat, N., I. Comyn-Wattiau, and J. Akoka (2015). "A taxonomy of evaluation methods for information systems artifacts". In: *Journal of Management Information Systems* 79.3, pp. 229–267. DOI: 10.1080/07421222.2015.1099390.

Rabiee, F. (2010). "Focus group interview and data analysis". In: *Proceedings of the Nutrition Society* 63.4, pp. 655–660. DOI: 10.1079/PNS2004399.

Ranki, M. (2019). "People, Process and Technology in CRM Implementation". Tampere University of Applied Sciences.

Reid, A. and M. Catterall (2014). "Hidden Data Quality Problems in CRM Implementation". In: *Developments in Marketing Science: Proceedings of the Academy of Marketing Science*, pp. 184–189. DOI: 10.1007/978-3-319-11779-9_67.

Robson, C. and K. McCartan (2016). *Real world research*. 4th ed. John Wiley Sons.

Rosing, M., J. Scheel, and A. Gill (2015). *Applying agile principles to BPM: the complete business process handbook*.

Runeson, P. and M. Höst (2007). "Guidelines for conducting and reporting case study research in software engineering". In: *Empir Software Eng* 14, 131–164. DOI: 10.1007/s10664-008-9102-8.

Sharma, S., D.P. Goyal, and R.K. Mittal (2010). "Imperative relationship between data quality and performance of datamining tools for CRM". In: *Int. J. Business Competition and Growth* 1.1, pp. 45–61. DOI: 10.1504/IJBCG.2010.032828.

Spruit, M. and V. van der Linden (2019). "BIDQI: The business impacts of data quality interdependencies model". In: *Department of Information and Applied Sciences*.

Spruit, M. and K. Pietzka (2015). "MD3M: The master data management maturity model". In: *Computers in Human Behavior* 51, pp. 1068–1076. DOI: http://dx.doi.org/10.1016/j.chb.2014.09.030.

Stone, M. et al. (2017). "How platforms are transforming customer information management". In: *The Bottom Line* 30.3, pp. 216–235. DOI: 10.1108/BL-08-2017-0024.

Sundararaman, a. and S.K. Venkatesan (2017). "Data Quality Improvement Through OODA Methodology". In: *Proceedings of MIT ICIQ*.

Thalheim, B. and Q. Wang (2013). "Data migration: A theoretical perspective". In: *Data & Knowledge Engineering* 87, pp. 260–278.

Tremblay, M.C., A.R. Hevner, and D.J. Berndt (2010). "The Use of Focus Groups in Design Science Research". In: pp. 1221–143. DOI: 10.1007/978-1-4419-5653-8_10.

Tseng, Yi-Hong and Ching-Torng Lin (2011). "Enhancing enterprise agility by deploying agile drivers, capabilities and providers". In: *Information Sciences* 181.17, pp. 3693–3708. ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2011.04.034. URL: https://www.sciencedirect.com/science/article/pii/S0020025511002088.

Tu, Y. and Z. Yang (2013). "An enhanced Customer Relationship Management classification framework with Partial Focus Feature Reduction". In: *Expert Systems with Applications* 40, pp. 2137–2146. DOI: 10.1016/j.eswa.2012.10.028.

Vancauwenbergh, S. (2019). "Data Quality Management". In: *Scientometrics Recent Advances*. IntechOpen, pp. 3–18. ISBN: 978-1-78985-613-2.

Vaziri, R., M. Mohsenzadeh, and J. Habibi (2016). "TBDQ: A pragmatic task-based method to data quality assessment and improvement". In: *PLoS ONE* 11.5, pp. 1–30.

Wang, R.Y. (1998). "A product perspective on total data quality management." In: *Communications of the ACM* 41.2, pp. 58–65.

Webster, J. and R. T. Watson (2002). "Analyzing the Past to Prepare for the Future : Writing a Literature Review". In: *MIS Quarterly* 26.2, pp. 13–23.

Weerd, I. van de (2009). "Advancing in Software Product Management: An Incremental Method Engineering Approach". In:

Weerd, I. van de and S. Brinkkemper (2008). *Meta-modeling for situational analysis and design methods*. Chap. III, pp. 35–54.

Weerd, I. van de, S. Brinkkemper, and J. Versendaal (2007). "Concepts for incremental method evolution: Empirical exploration and validation in requirements management". In: *Advanced information systems engineering*. Ed. by J. Krogstie, A. Opdahl, and G. Sindre. Vol. 4495. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 469–484. DOI: 10.1007/978-3-540-72988-4_33.

Wieringa, Roel.J. (2014). *Design Science Methodology: for Information Systems and Software Engineering*. Heidelberg: Springer. DOI: 10.1007/978-3-662-43839-8.

Wijnhoven, F. et al. (2007). "Total Data Quality Management: A Study of Bridging Rigor and Relevance". In: *ECIS 2007 Proceedings* 15.

Yin, R.K. (2013). *Case study research: Design and methods*. Sage publications.

Zahay, D., A. Krishen, and J.W. Peltier (2012). "Building the foundation for customer data quality in CRM system for financial services firms". In: *Journal of Database Marketing Customer Strategy Management* 19.1, pp. 5–16. DOI: 10.1057/dbm.2012.6.