



# Designing a Data Quality Management Framework for CRM Platform Delivery and Consultancy

Renee Albrecht<sup>1</sup> · Sietse Overbeek<sup>1</sup> · Inge van de Weerd<sup>1</sup>

Received: 29 November 2022 / Accepted: 28 July 2023  
© The Author(s) 2023

## Abstract

CRM platforms heavily depend on high-quality data, where poor-quality data can negatively influence its adoption. Additionally, these platforms are increasingly interconnected and complex to meet the growing needs of customers. Hence, delivery and consultancy of CRM platforms becomes highly complex. In this study, we propose a CRM data quality management framework that supports CRM delivery and consultancy firms to improve data quality management practices within their CRM projects. We develop the framework by extracting best practices for CRM data quality management by means of a literature study on data quality definition and measurement, data quality challenges, and data quality management methods. In a case study at an IT consultancy company, we investigate how CRM delivery and consultancy projects can benefit from the incorporation of data quality management practices. The results translate into a framework that provides a high-level overview of data quality management practices incorporated in CRM delivery and consultancy projects. It includes the following components: Client profiling, project definition, preparation, migration/integration, data quality definition, assessment, and improvement. The framework is validated by means of confirmatory focus groups and a questionnaire.

**Keywords** Consultancy · CRM · Data quality management · Delivery · Design science

## Introduction

Contemporary enterprises heavily depend on data, where data are seen as a strategic asset [42]. The International Data Corporation (IDC) states that during 2020, the amount of data that have been created, captured, and replicated across the world is more than 59 zettabytes, and over the next 5 years, the world will create more than three times the amount of data compared to the previous 5 years [31]. With an increasing volume, velocity, and variety of data, the issues

arising from errors in data and the organizational impact of these issues are amplified [35]. 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” [35, pp. 246–247]. Poor data quality (DQ) is one of the greatest challenges facing contemporary enterprises [18]. Simultaneously, enterprises struggle to address their data issues, while high-quality data are rather the exception than the rule [42]. 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 utilization of high-quality data is customer relationship management (CRM). Poor DQ and integration can negatively influence the adoption of CRM [16]. Additionally, survey data collected from about 300 organizations 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 to 20% of total revenue [29]. CRM is defined as “the core business strategy that aims to create and maintain profitable relationships with

---

This article is part of the topical collection “Advances on Enterprise Information Systems” guest edited by Michal Smialek, Slimane Hammoudi, Alexander Brodsky and Joaquim Filipe.

---

✉ Sietse Overbeek  
s.j.overbeek@uu.nl

Renee Albrecht  
reneealbrecht04@gmail.com

Inge van de Weerd  
i.vandeweerd@uu.nl

<sup>1</sup> Department of Information and Computing Sciences,  
Faculty of Science, Utrecht University, Princetonplein 5,  
3584 CC Utrecht, The Netherlands

customers, by designing and delivering superior value proposition” [12, 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 to a variety of customers. 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 [58].

The objective of this research is to provide a solution for the need for data of adequate quality in increasingly complex CRM platforms. The aim is to improve data quality within CRM platforms by designing a framework that supports data quality management in order to keep the quality of CRM data on an adequate level. This translates to the following research question: *How can a data quality management framework be designed to support CRM platform delivery and consultancy?*

This study has several contributions. First, the investigation of CRM and its D &C provides insights in what CRM D &C projects entail and which practices have to be taken into consideration. Second, the understanding of which types of data are utilized in contemporary CRM platforms serves as input on how a definition of CRM DQ can be established, as well as how CRM DQ can be measured. Third, the investigation of known challenges regarding DQ within CRM platforms and the potential solutions. In this section, we present our multivocal literature review, which contributes to the establishment of the CRM-DQMF. Furthermore, we elaborate on the CRM platforms provided by IT consultancy firms are elaborated on, as well as their D &C practices. ojects results in a list of criteria for DQ management in CRM D &C projects.

The following section introduces the research approach. Subsequently, we present the results of a literature review in “[Literature study results](#)”. A CRM DQ management framework is introduced in “[CRM DQ management framework design](#)”, followed by its validation in “[Validation](#)”. A conclusion and discussion of our research is elaborated in “[Validation results](#)”, as well as, an exploration of avenues for further research.

## Research Approach

In this study, we adopt a design science approach. Our aim is to investigate and design a framework for data quality management in CRM platform D &C. In line with Wieringa’s Design Science Framework [68], we follow an iterative set of problem-solving tasks according to the structure of the so-called design cycle.

The first task is *Problem Investigation*, where the design of the CRM DQ Management Framework (CRM-DQMF) is prepared by conducting exploratory research to understand the problem. To increase the robustness of the results of this research, methodological triangulation is adopted [33]. We carried out a literature review to identify, evaluate, and integrate findings of relevant high-quality studies that address the research problem. We determined the relevance of the literature by scope, objectives, methods, and conclusions [11]. To ensure a complete review, additional search methods that are used are forward and backward searching [37]. We also included gray literature to get insights on the state of the art concerning DQ management in CRM platforms in practice.

The second part of our problem investigation was carried out in the form of a single embedded case study at an IT consultancy firm [70]. The goal of this case study was to investigate the defined problem within its context. Its results are triangulated by including exploratory expert interviews with a total of 14 experts, as well as a documentation analysis including 15 relevant documents mainly existing of company documentation. The participating experts and researched documentation can be found in Tables 1 and 2, respectively.

The second task, *Treatment Design*, specifies requirements for the CRM-DQMF, which are extracted from insights of the triangulated collected data. Based on those requirements, the design of the CRM-DQMF is established. The results are presented by means of a Process-Deliverable Diagram (PDD) consisting of two integrated diagrams, namely a process diagram including all activities, and a deliverable diagram including the deliverables that result from the activities [63, 64]. We used an assembly-based method engineering approach that facilitates situational analysis and design methods. Since situational factors play

**Table 1** Participants expert interviews

ID	Role	Years of experience
P1	Consultant	4
P2	Senior analyst	2
P3	Senior director	14
P4	Group manager	15
P5	Senior consultant	13
P6	Senior consultant	5
P7	Senior analyst	4
P8	Manager	9
P9	Consultant	4
P10	Consultant	6
P11	Analyst	2
P12	Senior consultant	7
P13	Director	14
P14	Senior consultant	7

**Table 2** Researched documentation

ID	Purpose	Year
D1	CRM platform insights	2020
D2	Data migration guidelines	2020
D3	CRM (on-premises) to cloud migration guidelines	2020
D4	CRM platform adoption guidelines	2020
D5	Common data model	2020
D6	Report on the solution design #1 for a utility client	2020
D7	Report on the solution design #2 for a utility client	2020
D8	Data flows explanation within a solution for a utility client	2020
D9	Functional design datamart for a utility client	2020
D10	Education in CRM platform functionalities	2021
D11	Global policy in data management	2021
D12	Dedicated system for data quality checks in multiple systems	2011
D13	Data management roadmap proposal	2021
D14	Data quality analysis strategy at a product client	2019
D15	Design principles in a CRM project of an NGO client	2021

a key role when managing DQ, e.g., the CRM platform, the industry, and data processes, this is deemed as a suitable approach for the design of the CRM-DQMF.

In the last task, *Treatment Validation*, we validate the initial design of the CRM-DQMF by means of expert opinions extracted using confirmatory focus groups with a total of 6 experts [61]. Subsequently, we conducted an interactive questionnaire among the same group of experts to effectively validate the first design of the CRM-DQMF [53].

### Threats to Validity

The five types of validity as described by Johnson [32] are used to examine the validity of this research.

*Descriptive Validity* 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 the validation session were recorded to facilitate more accurate recalls of the researcher.

*Interpretive Validity* To ensure this type of validity, the researcher regularly incorporated participant feedback within the case study [32]. This is done through the utilization of the question type *interpreting questions* as proposed by Kvale [34] to inspect whether the interviewee's answer is interpreted correctly.

*Theoretical Validity* To achieve theoretical validity, fieldwork is incorporated in this research. This fieldwork consists of several elements: participation in training sessions which were facilitated to CRM D & C consultants;

close observation of a collaboration tool used by the CRM D & C community at the case company; attending presentations on the execution of specific CRM D & C projects; joining a day of scrum meetings between the case company and a client of the financial industry; and gain certificates which are required/recommended for CRM D & C employees.

*Internal Validity* Method triangulation is utilized to achieve internal validity [32, 33]. 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 [32]. Multiple expert interviews are conducted with participants from varying backgrounds. Subsequently, documentation from a variety of sources is examined for the documentation analysis.

*External Validity* The case study is performed at one organization. However, various literature sources are used for the design of the CRM-DQMF, and documentation utilized for the documentation analysis originated from two additional organizations. Furthermore, the abstractness and high-level approach of the CRM-DQMF increases its external validity, as it facilitates generalizability to CRM D & C projects varying in client, industry, and business goals.

### Literature Study Results

This section presents the results of our literature study on CRM platforms, CRM data, and data quality management. The findings of the study serve as input for the CRM-DQMF.

### CRM and 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 [40]. A CRM platform is a CRM software package that can be utilized by organizations to systematize their CRM [15], either on-premise or as a SaaS solution. Customers wishing to implement a CRM platform in the cloud are able to utilize its functionalities by uploading their data to the host's servers and interact with the data using a web browser [12].

For CRM to be successful, business processes, human factors, and technology need to be integrated [44]. 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 organization, require limited economic and human resources' effort, and be promptly deployed by specifically trained employees [15].

## CRM Platform Delivery and Consultancy

When implementing a CRM platform, people, processes, and technology are crucial components [51]. 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 [16]. 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 [12].

For the delivery of a new CRM platform, the data that will be utilized need to be migrated from the original CRM system's data sources to the new platform data sources [2, 60]. 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 [28]: 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 [1, 28]. However, those standard processes may not completely fit into the needs of the specific organization [10]. 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 ETL processes [60]: *Extract* the data from the legacy data sources; *Transform* the data to meet the requirements of the new CRM platform; and *Load* the data into the targeted data sources of the new platform. After migration, data gathering for the utilization 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 [58].

## CRM Data

This section provides an overview the data that are utilized in CRM platforms. We specifically focus on data types and data quality (DQ). Building on this, literature is reviewed to find methods for the measurement of the quality of CRM data. Subsequently, existing DQ challenges and potential solutions are examined.

### CRM Data Types

Data represent the foundation for generating CRM-related information, serving as a key enabler for efficient processes [46]. According to [40], there are different types of data that are used in CRM systems:

- *Customer/prospect contact data* represent the most basic data in 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 data* include values such as gender, marital stage, income, and ethnicity.
- *RFM/transactional data* include information such as the customer's last purchase date, frequency of purchase, or customer service activities. The data are required for RFM calculations to analyze customer value, as well as defining customer retention and lifetime value. It is behavioral data and can be classified as predictive.
- *Psychographic data* consist of information such as beliefs, opinions, lifestyles, values, or motives of customers. This information can be utilized for the customer relationship processes, as it is an added dimension to predictive modeling and denotes intended behavior.
- *Customer touchpoint data* include information captured via the internet, email click-throughs, service encounters, and telephone calls.
- *Personalization data* represent the capability to tailor communications to the individual customer. Its data include information that is targeted at specific customers.

These types of data can be classified into either structured, unstructured, or semi-structured data [12, 43]. Table 3

**Table 3** CRM data types

Data structure	Definition	Example
Structured data	A data structure is provided by a pre-defined model specifying how the data should be stored	Relational data tables with prospect data
Semi-structured data	Data that have a structure with some degree of flexibility	Web pages; Twitter data
Unstructured data	Textual or non-textual data without a pre-defined model; A generic sequence of symbols	Customer emails and phone calls

presents an explanation as well as examples of each of the structures.

### CRM Data Quality

CRM data quality can be defined as the fitness to serve a purpose in the context of CRM [65]. It is a multi-faceted construct, consisting of a set of DQ dimensions [46, 67]. The dimensions are highly context dependent and their relevancy can vary between organizations and types of data. Hence, the identification of the relevant quality dimensions builds the basis for the assessment of DQ and its possible improvements [14].

#### Data Quality Dimensions

There is an extensive amount of different DQ dimensions and definitions in literature, of which a most common set includes *accessibility*, *accuracy*, *completeness*, *consistency*, *integrity*, and *timeliness* [4, 8, 14, 46, 48, 67]. Based on this literature, we come to the following definitions:

- *Accessibility*: 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 organization obtains the name ‘John’ for a customer, but the actual known name is ‘Jonas’, then the data are not accurate.
- *Completeness*: 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 are complete when the middle name is missing, yet incomplete when the first name is missing.
- *Consistency*: The extent to which there exists contradiction in data. Semantic rules specify the correctness of data. For example, data are 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 pre-defined format, such as ‘dd/mm/yy’ for a customer’s date of birth.
- *Timeliness*: 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.

#### Measuring Data Quality

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 [14]. This is done using metrics, which 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 [30]. To make objective assessments, organizations should comply to certain principles and develop metrics specific to their needs [21, 48]. 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 [14]. Once the task of defining the accurate dimensions is completed, the following ways of combining DQ metrics can be formulated [48]:

- *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.
- *Min or Max operations* 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.
- *Weighted average* is an alternative to the min operation. When an organization 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 DQ dimensions mentioned previously, examples of metrics along with their explanations can be found in Table 4.

**Table 4** Dimensions and metrics

Dimension	Metric	Explanation
Accessibility	Max operations	The maximum value of two terms: 0 and 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 are no longer useful. Delivery refers to when the data are 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
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 [4]. 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 are received by the system and age refers to the age of the data when first received by the system

Finally, to conclude whether the data are of sufficient quality, thresholds to compare the calculated metrics to need to be defined [17]. Whenever a measurement exceeds the pre-defined threshold, the data can be concluded to be of insufficient quality. This threshold can differ per organization or moment in time. For example, the *completeness* for the value *last name* in a financial organization might be of more relevance than for an organization in the product industry. Each organization has to adhere to various internal as well as regulatory policies that might define DQ thresholds [17].

## Challenges

High-quality integrated customer data are the foundation of successful CRM projects. Most prominent problems regarding CRM relate to the quality of data and its integration [40]. Literature suggests various challenges and solutions regarding CRM DQ, of which an overview can be found in Table 5. These challenges serve as input for the DQ management framework for CRM.

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).

## Data Quality Management

This section reviews existing literature about DQ management methods in the context of CRM platforms. DQ management refers to quality-oriented data management, where

the management focuses on the the collection, organization, storage, processing, and presentation of data of high quality [46]. In the following sections, existing DQ management methods are explained, after which the applicability to CRM is elaborated on.

### Existing Data Quality Management Methods: A Review

Existing scientific literature on DQ management methods appears to be scarce [69] 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 [52], where DQ management is a continuous process. If 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 [56]), the organization recognizes the root causes of poor DQ, and the organization establishes improvement practices when DQ is found to be poor [57]. 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 [39].

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* [14], 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 "Challenges" is formulated [25, 57]. The criteria are listed below, along with the activities that should be identified within the method and the related solution(s).

**Table 5** 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 [55]. Different records in a CRM platform may require different levels of DQ defining optimal quality. A model that allows different levels of DQ to be set for different records in a CRM platform is proposed by [22]
C2	Inconsistency in data entry and storage	Creation of a framework for data standards and rules, as well as a process for the verification of data [47, 62]; Analysis of whether it is more beneficial to focus on process redesign or simplification rather than on the integration of an existing process [24]. For example, the way of collecting customer data can lead to increased customer DQ in CRM platforms [71]. 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 [47]; Automation of data collection; Continuous inspection, correction and cleaning [22]; utilization of error detection tools [40]
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 utilized by the new CRM solution. This can be done by using a data cleanup tool [27, 52]
C5	Inadequate integration of data sources	Usage of specialized tools or data models for the integration of data of different formats [47, 58]
C6	DQ deteriorating over time	(Automatic) Updates and linkage of data throughout the entire CRM platform [47]
C7	High costs for maintaining DQ	Cost reduction can be achieved by addressing key quality problems from the start of the CRM implementation project [22]
C8	Big data	Big data creates problems for conventional DM techniques used in CRM, because of its volume, velocity, and different data structures [43]. The proposed solution is big data analytics, for which cloud computing provides the infrastructure [3]. The SaaS firm should have a standardized framework for big data analysis in place, which could produce satisfactory results for the organization's CRM data [62]

- *Upfront Considerations:* Reconstruct the organizational environment. Define a data standardization that is organization-wide. This should include structured, semi-structured, and unstructured data.
- *DQ Assessment:* DQ is defined regarding the requirements of 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 6 [8, 14]. A more detailed description along with the phases of the individual methods can be found in the Appendix.

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

**Table 6** Data quality management methods

Acronym	Name	References
AIMQ	A methodology for information quality assessment	[36]
CDQ	Comprehensive methodology for data quality management	[7]
COLDQ	Cost-effect of low data quality (Loshin Methodology)	[38]
DQA	Data quality assessment	[48]
DQAF	Data quality assessment framework	[9]
DQPA	A data quality practical approach	[14, 19]
HDQM	A data quality methodology for heterogeneous data	[5]
HIQM	Hybrid information quality management	[13]
OODA DQ	The observe-orient-decide-act methodology for data quality	[59]
TBDQ	Task-based data quality method	[66]
TDQM	Total data quality management	[67]
TIQM	Total information quality management	[20]

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 meta-data containing quality scores of data sources.

Despite differences in the names and individual steps that are adopted by the distinct methods, a general case with a set of basic steps composed of three phases can be extracted: *state reconstruction*, *assessment (including DQ definition)*, and *improvement* [8]. 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 organization, 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 play 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 [14]. In

general, the assessment phase consists of the following set of basic steps [8]:

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 [30, 56, 57]. Eleven business impacts of poor DQ to be included in the business impact analysis are found in [56].
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-structured data implicitly. For TDQM and CDQ semi-structured data are 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



**Table 7** 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

dimensions can also vary per method. In Table 7, 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. The various adoptions of DQ dimensions within the distinct methods confirms the gap regarding an effective standardization of dimensions as mentioned in “Challenges”.

The DQ measurement strategies and techniques differ between the distinct methods. As mentioned in “CRM Data Quality”, 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 8.

#### Improvement

This phase defines the steps and strategies that need to be undertaken for reaching new DQ targets. Organizations

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 [8]:

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 [30]. A detailed and complete classification of costs is also available [6].
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 organization. The improvement strategies can be either *process-driven*, meaning the quality of data is improved

**Table 8** 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

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 organizational 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 emphasizes 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 to assess its value [14]. Most methods emphasize 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 9.

#### Data Quality Management Within a CRM Delivery and Consultancy Project

Concluding from literature, the utilization of a mix of the existing DQ management methods for the context of CRM should be applied, using the best of them [8, 14, 25, 46].

**Table 9** Strategies for deciding on 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
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

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, 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. The following list of activities is extracted and synthesized from literature to be included in a CRM-DQMF:

- Upfront considerations to define the CRM D & C and reconstruct the organizational 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. To 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.

## CRM DQ Management Framework Design

From the expert interviews as well as the documentation analysis as part of the case study, it 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. Experts that participated in interviews state that CRM D & C teams face the challenge of the existing client’s context, with its own constraints and levels of expertise regarding DQ and DQ management. DQ management practices are not standardized within CRM D & C projects at the IT company of the case study, and are applied and performed subjectively varying depending on the experience(s) of the expert. Additionally, experts stated that there is need for DQ management, as found challenges in CRM D & C projects are largely related to poor DQ. Experts mutually experienced that currently,

assessment and improvement of DQ is not done proactively, while they agree that this would benefit the D & C project. 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, from the start of the project.

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. From the findings in literature and the case study, the requirements for the CRM-DQMF as explained in the following section are extracted. This is followed by an elaboration on the individual components of the CRM-DQMF in “[Activities and deliverables](#)”.

## Requirements Specification

For the CRM-DQMF to assist CRM D & C teams effectively in the management of DQ, it should adhere to certain criteria that ensure the management of DQ in general, as well as the management of DQ in the context of CRM D & C projects. It includes *Modularity, DQ Management Plan, DQ Management Maturity Level, CRM D & C Client Context, Migration/Integration, Iteration, Business Impact Analysis, DQ Assessment, and DQ Improvement*.

### Modularity

Customers expect a CRM platform to be tailored for their organization specifically, with limited effort, and deployed promptly [15]. Insights as found in the case study indicate customer-centric and agile approaches for CRM D & C projects. Hence, to make DQ management in CRM D & C projects succeed, the CRM-DQMF is required to be designed in a modular fashion. Modularity refers to the uniqueness of every client and CRM D & C project, and therefore the need for uniqueness in DQ management application. 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. The modular visualization is facilitated by making use of situational method engineering as proposed by van de Weerd and Brinkkemper [63]. By the introduction of an activity which produces a CRM D & C project specific DQ management plan as explained in “[DQ Management](#)

Plan”, the remainder of the utilization of the CRM-DQMF is decided upon.

### DQ Management Plan

The CRM-DQMF facilitates the establishment of a unique DQ management plan at the start of any project. It describes the roles and responsibilities of the client and 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 matters that make a client and project unique. This includes a business case, as this formats the required DQ management practices. The business case comprises of the client’s budget, the client’s business goals of a CRM D & C project, and the scope of the project. The budget indicates to what extent the client will be able to pay for DQ management services. The business goals provide an indication of the need for DQ management. The scope of the project indicates which functionalities are required for reaching the business goals, thus provides for an indication of the extent to which DQ management is required. Additionally, this includes the impact of DQ on the business goals, and the current DQ management maturity level and goals of the client (see “[DQ management maturity level](#)”).

### DQ Management Maturity Level

Results of the experts interviews indicate that DQ, and therefore the need for its management within a CRM D & C project along with the role of the CRM D & C team, depends on the expertise of the client. P3 mentioned: “In the ideal case, organizations already have an authority in place that takes care of data quality matters. However, this varies per organization and industry. The interference of us depends on the arrangements with the client”. Additionally, documentation mentions various DQ Key Performance Indicators, such as the number of data elements with a definition. This indicates that there is need for the determination of the client’s DQ management maturity level, which determines the extent to which DQ management will be applied and by whom. On the one hand, the client might not have any knowledge on their own data, nor its quality, which might indicate that the data are not of sufficient quality for a CRM solution and the client does not have sufficient in-house DQ management expertise, meaning the expertise of the CRM D & C team is required. On the other hand, the client might already be in control of its data (and quality) across the organization, which means there is no need for the CRM D & C team to conduct or advise on any DQ management practices. Therefore, the current DQ management maturity level of the client as well as the goal DQ maturity level of the client play relevant roles for the establishment of the DQ management plan. To determine the DQ management maturity level of

the client, the maturity matrix by Spruit and Pietzka can be utilized [57]. The capabilities of this maturity matrix are confirmed by this research. Those capabilities read: *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*.

### CRM D & C Client Context

Taking upfront considerations into account is considered to be important for DQ management in CRM D & C projects [8]. This is confirmed by the case study in terms of the definition of a client context. Documentation showed the establishment of a so-called blueprint of the project, which is defined to create the framework for the CRM solution based on budget, goal, and scope. Experts mentioned this blueprint, as well as practices for the reconstruction of business processes of the client. The client context comprises information for a reconstruction of the organizational environment in regards to the CRM solution, which includes business processes, 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.

### Migration/Integration

CRM D & C projects comprise the process towards a CRM solution, which typically includes the migration and/or integration of data. P4 explained that “Business requirements are defined to decide which data are required for the solution. From this, a data mapping is realized to load the data to the new CRM solution correctly”. Migration and integration practices comprise the mapping of data of two different systems: a legacy system and the new CRM solution [2]. Therefore, the CRM-DQMF should contain guidance to integrate DQ management into migration and integration practices, including data mapping practices. The practices of the *Transform* phase of the ETL process which is required for data migration and integration practices as found in literature [60] are confirmed by experts and documentation of the case study. This includes the mapping of the data, as well as the assessment and improvement of DQ.

### Iteration

In terms of the CRM D & C project, iteration needs to take place to successfully establish migration and integration, as the case study indicates that the 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 [8, 14]. Experts mention phrases such as “I think data quality should be measured frequently in any case”, and in documentation as well as by experts the data lifecycle is mentioned, which consists of the creation, management, and destruction of data. This indicates that DQ should be managed up until its destruction. 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 the lifecycle of all concerned data within the scope of the CRM D & C solution ceases to exist.

### Business Impact Analysis

The business impact of poor DQ needs to be analyzed, as this defines the data elements that are critical for the client’s business goals and thus require DQ assessment and potentially improvement practices [8, 14, 30]. This is also referred to as a *top down* or *demand driven* approach. 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 are recommended to be included within the impact analysis, as they are found to be applicable for a variety of industries [56]. Furthermore, the case study indicates agile and customer-centric project approaches. By not exclusively including monetary impact, the business impacts support an agile project approach, as agile values an emphasis on the quality, the flexibility and the customer-centricity of services [54] and cost efficiency is not at the center of attention in an agile project approach [26]. 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 [6].

### DQ Assessment

Within DQ management, DQ assessment is found to be a critical part and should take place frequently [8, 14]. Therefore, the CRM-DQMF should include guidance in assessment practices. This includes the definition and measurement of DQ and reporting on potential DQ issues. DQ can have different definitions, which is highly context dependent [46, 67]. Each client and project is unique, meaning DQ is required to be defined for every project. A DQ definition is expressed in terms of DQ dimensions and DQ thresholds [14, 46, 67]. The unstructured, semi-structured, and structured data used by CRM need to be taken into account when defining the appropriate DQ definition [40, 71]. For each DQ definition, DQ metrics need to be defined to be able

to quantify DQ [14, 48]. Additionally, for DQ assessment and improvement to succeed, the assignment of data roles is required [8]. From the case study can be concluded that, without someone responsible for the data, its quality will not be managed.

### DQ Improvement

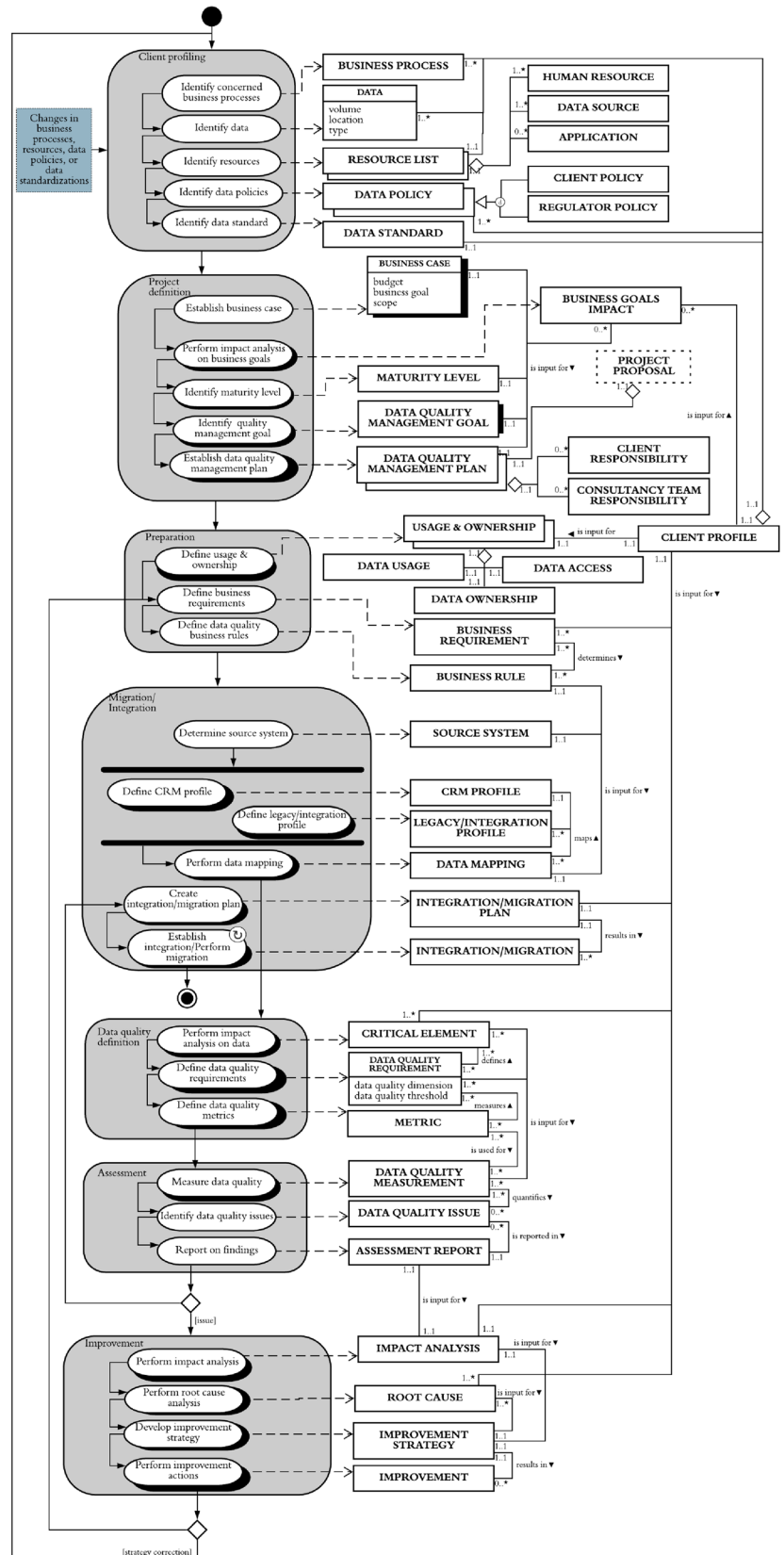
When data are found to be of insufficient quality, the CRM-DQMF should offer guidance in establishing an improvement strategy [8, 14]. For optimal DQ, an organization needs to be aware of different reasons for poor DQ and where they are existent within the organization, hence the root causes of DQ issues need to be analyzed [8, 14, 57]. The whereabouts of the weak spots should be known, as well as the reason(s) for the existence of weak spots, serving as input for the establishment of an improvement strategy. Direct and indirect costs of DQ are compared by making use of a cost evaluation to support a decision making process for its development of an improvement strategy [57]. These costs include the costs of the business impacts of the DQ issues, as well as the costs of potential improvement practices.

### Activities and Deliverables

The entire CRM-DQMF comprises seven phases, as can be seen in Fig. 1.

*Client profiling* is performed to reconstruct the client’s organizational 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. 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 utilization of the remainder of the CRM-DQMF. *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. Once the migration or integration is established, the data lifecycle does not come to an end, hence DQ is still required to be managed. 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 framework is reconstructed

Fig. 1 The CRM-DQMF



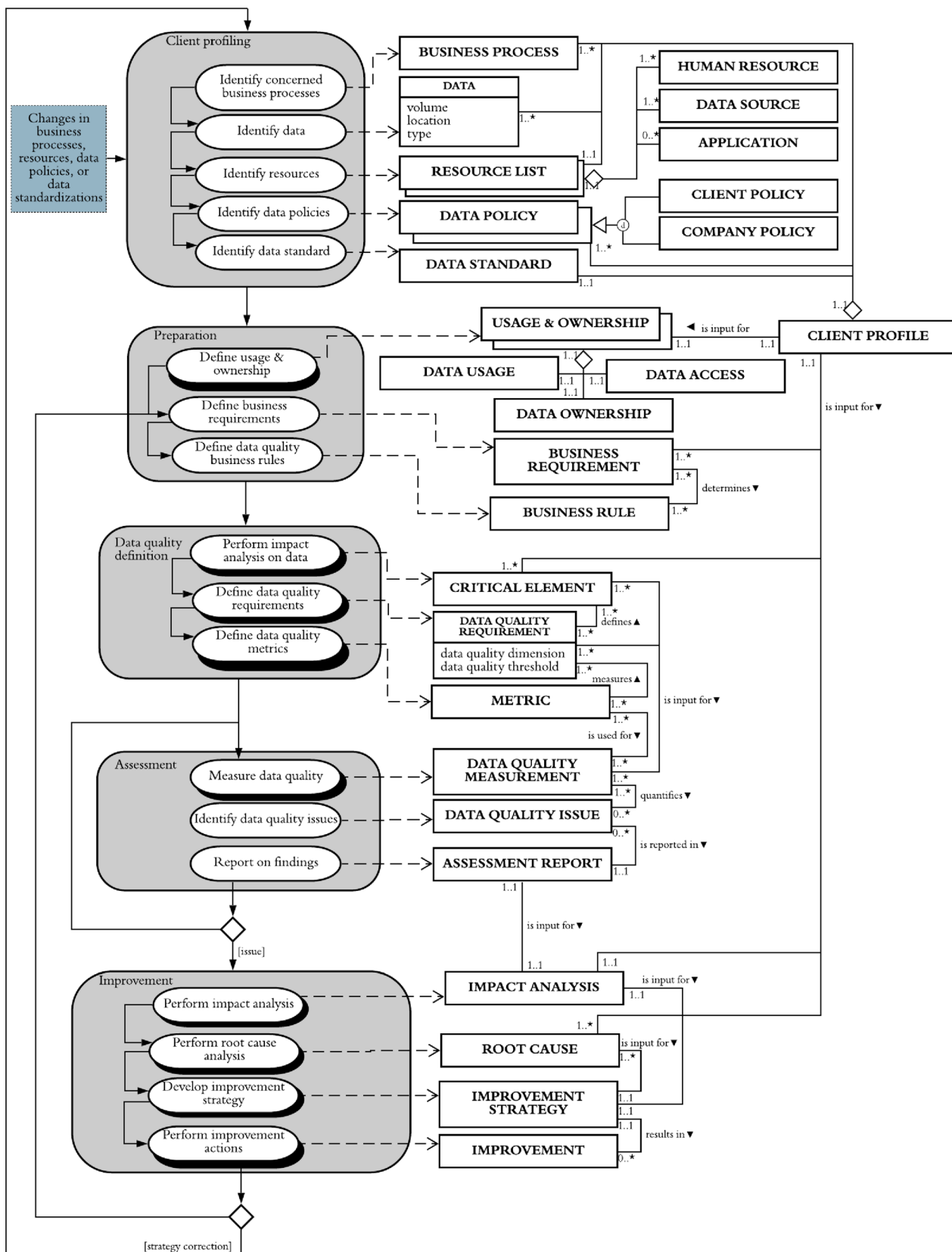


Fig. 2 The CRM-DQMF after migration/integration is established

as can be seen in Fig. 2. The distinct phases along with their activities are elaborated on below.

### Client Profiling

To understand and reconstruct the client's organizational environment with regards to the CRM solution, a client profile is established containing information on the data, business processes, resources, data policies, and data standards. Below, the different activities that gather the required information are explained.

*Identify Data* The concerned data are identified, so it is known which data should be subject to DQ management practices. According to Zahay et al. [71] and Missi et al. [40] there are different types of data within CRM, namely customer/prospect contact data; demographic; RFM/transactional data; psychographic data; customer touchpoint data; and personalization data. For example, an e-commerce organization makes use of its customer's e-mail addresses to send the confirmation of an order, a customer's gender to show a personalized webshop, and the last purchase date to use of the analysis of customer value. The different types of data can be classified into either structured, unstructured, or semi-structured data [12, 43]. The volume of the data is required to appropriately define a migration or integration strategy and to properly indicate the magnitude of potential DQ issues. The location of the data is 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 CRM solution are identified. The business processes create, use, move, or modify the concerned data, and form a technical and business process landscape indicating the whereabouts and purposes of the data. To give another example for the e-commerce organization, when a customer clicks on a link within a marketing email and places an order, contact data might be updated with a telephone number, touchpoint data are updated with email click-throughs, and the customer's last purchase date is updated.

*Identify Resources* The resources of the data are identified. This includes human resources, such as employees that enter the data, data sources that produce the data, and applications that utilize, 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 are identified, as well as regulatory 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 GDPR. The data policies are input for the definition of DQ requirements.

*Identify Data Standards* 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 continuous reviewing and refinement.

### 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. 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 are required to be of high quality, and to what extent DQ management is of relevance. An example of a business goal that could apply for an e-commerce organization can be found in Table 10. 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

**Table 10** Examples of different concepts used in the preparation phase

Term	Example
Business goal	Improve the marketing strategy by implementing this in a CRM system
Business requirement	Send an e-mail to a certain sample of end-customers once a month
Business rule	Two contacts cannot have the same e-mail address; e-mail addresses should contain an "@"
DQ requirement	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)



on DQ to deliver client experience within the constraints of costs and technique, or on the budget from the client's side.

*Perform Impact Analysis on Business Goals* By defining the impact of poor DQ on the business goals, as well as the way that high DQ will enable the business goals the importance of DQ management is 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 can 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 less expensive.

*Identify Maturity Level* The DQ management maturity level of the client is 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, and its relevance is supported by Spruit and Pietzka [57].

*Identify DQ 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 regard to the CRM D & C project. This is used as input for the establishment of the DQ management plan.

*Establish DQ Management Plan* Formatted by the business case, and using the input of the impact analysis, the current maturity level, and the goal maturity level of the client, the CRM D & C team develops a DQ management plan together with the client. 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. The plan determines the remainder of the utilization of the CRM-DQMF.

## Preparation

The preparation phase gathers the required information for the data mapping and the assessment.

*Define Usage and Ownership* For all data, the usage, ownership, and access are 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 requirements are 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. In Table 10, we present an example of a business requirement. To clarify the differences and relations between the various concepts, definitions of business goal, business rule, and DQ requirement are included as well.

*Define DQ Business Rules* Business rules are defined and refined, describing expectations about 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 are 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 *e-mail 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 *e-mail address* to be mandatory.

## 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. 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 [28]: on-premise to cloud; cloud to cloud; cloud to on-premise; or on-premise to on-premise. Source systems for integration may include third-party software, organizational database systems, transactions, and networked touch-points, such as social media or e-mail [16]. Once the DQ is assessed and, when required, improved, a migration and/or integration plan is created and executed. This is done after (some iterations of) assessment and improvement practices, since ideally potential issues regarding DQ are resolved before migration or integration is established to prevent more significant problems. 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 [63]. It indicates that migration practices can be performed multiple times as dry runs. When the issues as discovered by a dry run are eliminated, the migration will be established either

again as a dry run, or in production. After the migration has 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. When DQ issues occur, improvement practices are implemented first.

## DQ Definition and Assessment

The input for the DQ assessment phase is the gathered knowledge of the previous phases. The output is a DQ report.

*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 prioritized list of data elements which can be used by the team to focus their work efforts. For example, the presence of sufficient e-mail addresses might be critical for e-commerce organizations seeking to improve their marketing strategy.

*Define DQ Requirements* DQ is defined by means of DQ dimensions and DQ thresholds. DQ dimensions enable the characterization of rules (e.g., e-mail address must be populated) and findings (e.g., e-mail 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.

*Define DQ 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 *e-mail address* to be mandatory, which translates into the DQ dimension *completeness*. The metric that can be used to measure the completeness of the field *e-mail 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.

*Measure DQ* DQ is measured either subjectively or objectively [14]. The metrics are used to quantify the measurements. The output is the quantified measurements of DQ.

*Identify DQ Issues* Based on the measurements and DQ business rules, DQ issues are identified. DQ issues are identified by setting status indicators for all data in terms of its dimension(s) and thresholds. For example, the status indicator of the dimension of completeness for the field *e-mail 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. When DQ issues occur, improvement practices will take place.

## 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 organizational environment.

*Perform Impact Analysis* The identified DQ issues are quantified and prioritized based on business impact. Business impacts include monetary costs of 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.

*Perform Root Cause Analysis* Ideally, the DQ issues are remediated at their root cause [8]. 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.

*Develop Improvement Strategy* Based on the impact analysis, an improvement strategy is developed, evaluating the costs of the issue against the costs of the required improvement actions. 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 a quantification of the business impacts.

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

## Validation

To validate the CRM-DQMF, we drafted a validation model of our framework. A validation model, or design theory, consists of a description of the properties of the artifact and the interaction with the problem context [68]. The discussions that form the design theory are facilitated by the use of confirmatory focus groups [61] with a total of six experts (see

**Table 11** Participants validation

ID	Role	Years of experience
P1	Consultant	4
P2	Senior analyst	2
P3	Manager	9
P4	Senior consultant	4
P5	Senior consultant	7
P6	Senior analyst	2

Table 11. Additionally, to observe and measure how well the CRM-DQMF supports DQ management in CRM D & C projects, the evaluation criteria of [41] and [49] are utilized. The constructs of desired qualities (*Perceived Alignment with CRM D & C* and *Perceived Effectiveness*, *Perceived Ease of Use*, *Perceived Usefulness*, and *Perceived Completeness* are measured by means of an interactive questionnaire, of which the results on a 5-point Likert scale can be found in Table 13.

### 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 artifact [61]. 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 it might generate ideas based on the input of others [50]. An overview of participants can be found in Table 11. Each participant is an employee working on CRM D & C projects.

### Validation Constructs and Statements

The definitions of the distinct constructs on which the validation of the framework is based are provided in the following list.

- *Perceived alignment with CRM D & C business* is defined as the perceived congruence of the framework with organizations and their strategy [49]. 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 framework.
- *Perceived effectiveness* is the perceived degree to which the framework achieves its objectives in a real situation [41, 49].

- *Perceived completeness* is the perceived degree to which the structure of the artifact contains all necessary elements and relationships between elements [49].
- *Perceived ease of use* is the perceived degree to which experts believe that the use of the framework is free of effort [41].
- *Perceived usefulness* is the perceived degree to which experts believe that the framework is effective in achieving the intended objectives of the method [41].
- *Intention to use* is the degree to which experts have an intention to use the framework [41].

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 12, an overview of all statements can be found.

### Validation Session

The validation session is performed by making use of an online setting in Microsoft Teams. The validation session starts with an explanation of the framework 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 are 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 framework, where they can ask questions when required. Afterwards, the participants work on the case where they apply the framework 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 “[Validation constructs and statements](#)”, which is made available

**Table 12** Constructs and statements

Constructs	Statements
Alignment with CRM D &C Business	I do think that the data quality management framework is consistent with an organization's CRM delivery and consultancy needs
	To my experience, the framework is not congruent/compatible with an organization's CRM delivery and 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 and consultancy projects
	Overall, I do think that the framework is difficult to use within CRM delivery and consultancy projects
Perceived usefulness	Overall, I do think that the framework is useful for CRM delivery and 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 and consultancy projects
	I would not use this framework for data quality management within CRM delivery and consultancy projects

**Table 13** Participants validation

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

through Qualtrics (see: <https://www.qualtrics.com/support/survey-platform/survey-module/survey-module-overview/>) to every participant individually. During the confirmatory focus group the statements are discussed one by one by the participants.

## Validation Results

*Project Definition:* When discussing the *Project definition* phase of the framework, a question that was raised amongst the participants of the validation session 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 session 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 session also confirmed

findings of the case study, 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 of a CRM solution as possible. Therefore, it can be concluded that for an 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 framework.

However, in the second and third validation sessions, 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 do not 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 and Preparation:* The validation session concluded that the *Client profiling* and *Preparation* phases are part of every CRM D &C project, and they agreed on the sub-activities 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 the validation session 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, the validation session 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 and Assessment:* By each validation session, the definition and assessment practices of DQ as introduced by the framework 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 this will be maintained. Currently, this is where mistakes are made”.

*Improvement:* The improvement practices as introduced by the framework were confirmed to be applicable, and were also recognized 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.

## Questionnaire Results

The results of the questionnaire can be seen in Table 13, 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 framework 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 framework 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 framework 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 framework too high level, and would require more guidance for understanding how to perform the individual steps of the framework. To quote one participant: “I do not think it is easy to use, but I also do not think it is 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 with a score of 4.5. Participants agreed on the fact that the framework creates awareness on what DQ management encompasses for CRM D &C projects, from which the participants concluded the framework 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 framework 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 framework 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 framework 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.

## 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 “[Validation results](#)”, the establishment of a DQ management plan 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 within 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 framework could be discarded, which also eliminates unprofitable efforts.

*Awareness:* The validation session 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 [17]. 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 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 session. The participants argue that they would like the framework 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 framework 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 the validation session, it can be concluded that the contemporary design of the framework is too high level to put into practice as it is. Participants agreed that

the framework 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.

## Conclusion and Discussion

Our framework 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 can be used to plan on opportunities for the incorporation of DQ management in CRM D & C projects. It 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; and
- *Improvement* to define an improvement strategy and improve DQ.

In contrast to existing approaches for DQ, such as TDQM [20], DQA [48], and OODA DQ [59], our approach is not a general data quality management approach, but specifically tailored for the CRM domain. We achieved this using best practices for CRM DQ management from literature and practice.

Although DQ management is an important topic CRM literature, no comprehensive methods or approaches for DQ management in CRM have been proposed. For example, [23, 71] conducted useful research into the optimization and cost-effectiveness of DQ management for CRM, but do not prescribe the steps necessary for good DQ management. Other authors, such as [25] evaluate the applicability of existing DQ management methods for CRM, and suggest which is most suitable, but do not provide a specific approach tailored to CRM.

Our framework, the CRM-DQMF is a tool to support consultants to improve data quality management in their project.

It is tailored to the CRM practice and it provides a comprehensive overview of activities and deliverables.

## Limitations

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. 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. This contributes to the formulation of some question types, such as follow-up and probing questions [34], 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 session 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 [45].

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 the case company. 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.

## Future Research

This research leaves multiple opportunities for further research. First, improvement opportunities can be found in *perceived ease of use* and *perceived completeness* based on the validation. 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 utilization 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.

Second, further empirical validation of this research is required. This research solely evaluates the design through a validation session 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 to include additional perspectives next to those that are utilized 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 organization. During the case study and validation session, 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. Fourth, the implementation of the CRM-DQMF might be too expensive as it is. 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.

Lastly, 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

- Data Quality Management Methods, see: <https://osf.io/xkquj/>.
- Interview Protocol, see: <https://osf.io/3c9ys/>.
- CRM-DQMF Design, see: <https://osf.io/375up/>.

## Declarations

**Conflict of Interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahmad N, Naveed Q, Hoda N. Strategy and procedures for migration to the cloud computing. In: International conference on engineering technologies & applied sciences. 2018. p. 5.
- Ali H, Moawad R, Hosni A. A cloud interoperability broker (cib) for data migration in saas. *Future Comput Inform J*. 2017. <https://doi.org/10.1016/j.fcij.2017.03.001>.
- Balachandran B, Prasad S. Challenges and benefits of deploying big data analytics in the cloud for business intelligence. *Int Conf Knowl Based Intell Inf Eng Syst*. 2017;12:1112–22.
- Ballou D, Wang R, Pazer H, Tayi G. Modeling information manufacturing systems to determine information product quality. *Manage Sci*. 1998;44(4):462–84.
- Batini C, Barone D, Cabitza F, Grega S. A data quality methodology for heterogeneous data. *Int J Database Manag Syst*. 2011;3(11):60–79.
- Batini C, Barone D, Mastrella M, Maurino A, Ruffini C. A framework and a methodology for data quality assessment and monitoring. In: Proceedings of the 12th international conference on information quality. 2007.
- Batini C, Cabitza F, Cappiello C, Francalanci C, di Milano P. A comprehensive data quality methodology for web and structured data. In: Proceedings of the 1st international conference on digital information management. 2006. p. 448–56.
- Batini C, Cappiello C, Francalanci C, Maurino A. Methodologies for data quality assessment and improvement. *ACM Comput Surv*. 2009. <https://doi.org/10.1145/1541880.1541883>.
- Batini C, Rula A, Scannapieco M, Viscusi G. Data quality to big data quality. *Database Manag*. 2015;26(1):60–82.
- Birthare S, Sharma R. Study on migration of on-premise erp to saas product. *Int Res J Mod Eng Technol Sci*. 2020;2(12).
- Budgen D, Brereton P. Performing systematic literature reviews in software engineering. In: ICSE '06 Proceedings of the 28th international conference on software engineering. 2006. <https://doi.org/10.1145/1134285.1134500>.
- Buttle F, Maklan S. Customer relationship management: concepts and technologies, 4th edn. Routledge; 2019.
- Cappiello D, Ficiaro P, Pernici B. Hiqm: a methodology for information quality monitoring, measurement, and improvement. In: Advances in conceptual modeling theory and practice. Springer; 2006. p. 339–51.
- Cichy C, Rass S. An overview of data quality frameworks. *IEEE Access*. 2019;7:2169–3536. <https://doi.org/10.1109/ACCESS.2019.2899751>.
- Cricelli L, Famulari F, Greco M, Grimaldi M. Searching for the one: customer relationship management software selection. *J Multi-Criteria Decis Anal*. 2019. <https://doi.org/10.1002/mcda.1687>.
- Cruz-Jesus F, Pinheiro A, Tiago O. Understanding crm adoption stages: empirical analysis building on the toe framework. *Comput Ind*. 2019;109:1–13.
- DAMA I. DAMA-DBoK: data management body of knowledge, vol. 2. Technics Publications; 2017.
- Davenport T, Harris J. Competing on analytics: updated, with a new introduction: the new science of winning. Harvard Business Review Press; 2017.
- del Pilar Angeles M, García-Ugalde F. A data quality practical approach. *Int J Adv Softw*. 2009;2(2):259–74.
- English L. Improving data warehouse and business information quality. New York: Wiley; 1999.
- Erceg V, Šereš L, Zoranović T. Data quality assessment. In: International conference on applied internet and information technologies. 2019. p. 112–16.
- Even A, Shankaranarayanan G, Berger P. Evaluating a model for cost-effective data quality management in a real-world crm setting. *Decis Support Syst*. 2010;50:152–63.
- Even A, Shankaranarayanan G, Berger PD. Evaluating a model for cost-effective data quality management in a real-world crm setting. *Decis Support Syst*. 2010;50(1):152–63.
- Foss B, Henderson I, Murray D, Stone M. Managing the quality and completeness of customer data. *J Database Mark*. 2002;10:139–58.
- Francisco M, Alves-Souza S, Campos E, De Souza L. Total data quality management and total information quality management applied to customer relationship management. In: ICIME 2017: Proceedings of the 9th international conference on information management and engineering. 2017; p. 40–45. <https://doi.org/10.1145/3149572.3149575>.
- Gill A, Henderson-Sellers B. Measuring agility and adaptability of agile methods: a 4 dimensional analytical tool. In: The IADIS international conference on applied computing. 2006.
- Goldenberg J. CRM in real time: empowering customer relationships, vol. 18. Medford: Information Today; 2008.
- Hai H, Sakoda S. SaaS and integration best practices. *Fujitsu Sci Technol J*. 2009;45(3):257–64.
- Hanson G. Tate of crm data management 2020 report. 2020.
- Heinrich B, Hristova D, Klier M, Schiller A, Szubartowicz M. Requirements for data quality metrics. *J Data Inf Qual*. 2018;9(2):45–61. <https://doi.org/10.1145/3148238>.
- IDC. Global datasphere. IDC. 2020.
- Johnson R. Examining the validity structure of qualitative research. *Education*. 1997;118(2):282–92.
- Kaplan B, Maxwell J. Qualitative research methods for evaluating computer information systems. Springer; 2005. [https://doi.org/10.1007/0-387-30329-4\\_2](https://doi.org/10.1007/0-387-30329-4_2).
- Kvale S. Interviews: an introduction to qualitative research interviewing. Thousand Oaks; 1996.
- Laney DB. Infonomics: how to monetize, manage, and measure information as an asset for competitive advantage. Taylor & Francis Group; 2018.
- Lee Y, String D, Kahn B, Wang R. Aimq: a methodology for information quality assessment. *Inf Manag Horiz*. 2002;40(2):133–46.
- Levy Y, Ellis T. A systems approach to conduct an effective literature review in support of information systems research. *Inf Sci J*. 2006;9:181–212. <https://doi.org/10.1049/cp.2009.0961>.
- Loshin D. Enterprise knowledge management—the data quality approach. In: Series in Data Management Systems. Morgan Kaufmann; 2004. p. 448–56.



39. Marsh R. Drowning in dirty data? It's time to sink or swim: a four-stage methodology for total data quality management. *J Database Mark Cust Strategy Manag.* 2005;12(2):105–12.
40. Missi F, Alshawi S, Fitzgerald G. Why crm efforts fail? A study of the impact of data quality and data integration. In: Hawaii international conference on system sciences. 2005. p. 38. <https://doi.org/10.1109/HICSS.2005.695>.
41. Moody D. The method evaluation model: a theoretical model for validating information systems design methods. In: ECIS 2003 proceedings. 2003. p. 79.
42. Nagle T, Redman T, Sammon D. Assessing data quality: a managerial call to action. *Bus Horiz.* 2020;63:325–37. <https://doi.org/10.1016/j.bushor.2020.01.006>.
43. Nam D, Lee J, Lee H. Business analytics use in crm: a nomological net from it competence to crm performance. *Int J Inf Manage.* 2019;45:233–45.
44. Negahban A, Kim D, Kim C. Unleashing the power of mcrm: investigating antecedents of mobile crm values from managers viewpoint. *Int J Hum Comput Interact.* 2016;32(10):747–64. <https://doi.org/10.1080/10447318.2016.1189653>.
45. Nyumba O, Wilson K, Derrick C, Mukherjee N. The use of focus group discussion methodology: insights from two decades of application in conservation. *Methods Ecol Evol.* 2018. <https://doi.org/10.1111/2041-210X.12860>.
46. Otto B, Wende K, Schmidt A, Osl P. Towards a framework for corporate data quality management. In: Australasian conference on information systems. 2007. p. 916–26.
47. Petrović M. Data quality in customer relationship management (crm)—literature review. *Strateg Manag.* 2020;25(2):40–7. <https://doi.org/10.5937/StraMan2002040P>.
48. Pipino L, Lee Y, Wang R. Data quality assessment. *Commun ACM.* 2002. <https://doi.org/10.1145/505248.506010>.
49. Prat N, Comyn-Wattiau I, Akoka J. A taxonomy of evaluation methods for information systems artifacts. *J Manag Inf Syst.* 2015;79(3):229–67. <https://doi.org/10.1080/07421222.2015.1099390>.
50. Rabiee F. Focus group interview and data analysis. *Proc Nutr Soc.* 2010;63(4):655–60. <https://doi.org/10.1079/PNS2004399>.
51. Ranki M. People, process and technology in crm implementation. Master's thesis, Tampere University of Applied Sciences. 2019.
52. Reid A, Catterall M. Hidden data quality problems in crm implementation. In: Developments in marketing science: proceedings of the academy of marketing science. 2014. p. 184–89. [https://doi.org/10.1007/978-3-319-11779-9\\_67](https://doi.org/10.1007/978-3-319-11779-9_67).
53. Robson C, McCartan K. Real world research. 4th ed. New York: Wiley; 2016.
54. Rosing M, Gill A. Applying agile principles to bpm. In: The complete business process handbook, vol. 1. Elsevier-Morgan Kaufman; 2015. p. 553–77. <https://doi.org/10.1016/B978-0-12-799959-3.00027-6>.
55. Sharma S, Goyal D, Mittal R. Imperative relationship between data quality and performance of datamining tools for crm. *Int J Bus Compet Growth.* 2010;1(1):45–61. <https://doi.org/10.1504/IJBCG.2010.032828>.
56. Spruit M, Linden V. Bidqi: the business impacts of data quality interdependencies model. Department of Information and Applied Sciences: Utrecht University; 2019.
57. Spruit M, Pietzka K. Md3m: the master data management maturity model. *Comput Hum Behav.* 2015;51:1068–76. <https://doi.org/10.1016/j.chb.2014.09.030>.
58. Stone M, Aravopoulou E, Gerardi G, Todeva E, Weinzierl L. How platforms are transforming customer information management. *Bottom Line.* 2017;30(3):216–35. <https://doi.org/10.1108/BL-08-2017-0024>.
59. Sundararaman A, Venkatesan S. Data quality improvement through ooda methodology. In: Proceedings of MIT ICIQ. 2017.
60. Thalheim B, Wang Q. Data migration: a theoretical perspective. *Data Knowl Eng.* 2013;87:260–78.
61. Tremblay M, Hevner A, Berndt D. The use of focus groups in design science research. Springer. 2010. [https://doi.org/10.1007/978-1-4419-5653-8\\_10](https://doi.org/10.1007/978-1-4419-5653-8_10).
62. Tu Y, Yang Z. An enhanced customer relationship management classification framework with partial focus feature reduction. *Expert Syst Appl.* 2013;40:2137–46. <https://doi.org/10.1016/j.eswa.2012.10.028>.
63. van de Weerd I. Advancing in software product management: an incremental method engineering approach. PhD thesis, Utrecht University. 2009.
64. van de Weerd I, Brinkkemper S. Meta-modeling for situational analysis and design methods. In: Handbook of research on modern systems analysis and design technologies and applications, chapter III. IGI Global; 2008. p. 35–54. <https://doi.org/10.4018/978-1-59904-887-1.ch003>.
65. Vancauwenbergh S. Data quality management. In: *Scientometrics Recent Advances*. IntechOpen; 2019. p. 3–18.
66. Vaziri R, Mohsenzadeh M, Habibi J. Tbdq: a pragmatic task-based method to data quality assessment and improvement. *PLoS One.* 2016;11(5):1–30.
67. Wang R. A product perspective on total data quality management. *Commun ACM.* 1998;41(2):58–65.
68. Wieringa R. Design science methodology: for information systems and software engineering. Heidelberg: Springer; 2014. <https://doi.org/10.1007/978-3-662-43839-8>.
69. Wijnhoven F, Boelens R, Middel R, Louissen K. Total data quality management: a study of bridging rigor and relevance. In: ECIS 2007 proceedings. 2007. p. 15.
70. Yin RK. Case study research: design and methods. 3rd ed. Thousand Oaks: SAGE Publications; 2003.
71. Zahay D, Krishen A, Peltier J. Building the foundation for customer data quality in crm system for financial services firms. *J Database Mark Custom Strategy Manag.* 2012;19(1):5–16. <https://doi.org/10.1057/dbm.2012.6>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.