A Capability Reference Model for Strategic Data Management

D I S S E R T A T I O N
of the University of St. Gallen,
School of Management,
Economics, Law, Social Sciences
and International Affairs
to obtain the title of
Doctor of Philosophy in Management

submitted by

Tobias Pentek

from

Germany

Approved on the application of

Prof. Dr. Boris Otto

and

Prof. Dr. Klaus Möller

Dissertation no. 5009

Difo-Druck GmbH, Untersiemau 2020

The University of St. Gallen, School of Management, Economics, Law, Social Sciences and International Affairs hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St. Gallen, May 18, 2020

The President:

Prof. Dr. Bernhard Ehrenzeller

Acknowledgments

To Julia, Emma, and Henri, for their love, support, understanding, and patience.

Duisburg, May 2020 *Tobias Pentek*

Zusammenfassung

Im Kontext der digitalen Transformation und der steigenden Bedeutung datengetriebener Geschäftsmodelle sind Daten als strategische Unternehmensressource von grosser Relevanz. Das Management dieser strategischen Ressource wird als Schlüsselqualifikation von Unternehmen gesehen. Sowohl die wissenschaftliche Gemeinschaft als auch Praktiker betrachten daher das Datenmanagement mit zunehmender Aufmerksamkeit. Allerdings fehlt ein ordnender Rahmen, der die relevanten Gestaltungsbereiche und deren Zusammenspiel aufzeigt, um die notwendigen Datenmanagementfähigkeiten zu gestalten, und der dabei die Anforderungen datengetriebener Unternehmen berücksichtigt.

Die vorliegende Dissertation schliesst diese Lücke, indem sie mit dem Data Excellence Model ein Referenzmodell für die Gestaltung von Datenmanagementfähigkeiten vorstellt. Als Design-Artefakt liefert das Modell eine «Blaupause» für den Auf- und Ausbau von Datenmanagement. Diese Dissertationsschrift erläutert den Gestaltungsprozess, beschreibt das daraus resultierende Artefakt und demonstriert seine Anwendbarkeit und praktische Nützlichkeit anhand mehrerer Fallstudien.

Summary

In the context of digital transformation and the increasing importance of data-driven business models, data has turned into an important corporate resource of strategic relevance. Both the scientific and the practitioners' community consider the management of data as a key capability. Despite the growing importance of data management, however, a framework for designing and structuring strategic data management in data-driven enterprises does not exist yet.

The doctoral dissertation bridges this gap by introducing the Data Excellence Model as a capability reference model for strategic data management. The model serves as a "blueprint" for implementing, developing, and assessing data management capabilities. The dissertation describes the model design process, presents the resulting artifact, and demonstrates its applicability and practical utility in several case studies.

Content Overview

Table of Contents

List of Abbreviations

AI	Artificial Intelligence
AR	Action Research
ADR	Action Design Research
B2B	Business-to-business
B2C	Business-to-consumer
BI	Business Intelligence
BIE	Building, Intervention, and Evaluation
CC CDQ	Competence Center Corporate Data Quality
CDO	Chief Data Officer
CEO	Chief Executive Officer
CFO	Chief Financial Officer
CDM	Corporate Data Management
CDQM	Corporate Data Quality Management
cf.	compare (Latin: confer)
CMM	Capability Maturity Model
CMS	Content Management System
DAMA	Data Management Association
DQM	Data Quality Management
DSR	Design Science Research
DXM	Data Excellence Model
EAD	Enterprise Data & Analytics
EDQM	Enterprise Data Quality Management
EFPIA	European Federation of Pharmaceutical Industries and Associations
e.g.	for example (Latin: exempli gratia)
EFQM	European Foundation for Quality Management
EMA	European Medicines Agency
et seq.	and the following page (Latin: et sequēns)
et seqq.	and the following pages (Latin: et sequentēs)
EU	European Union
FMCG	Fast-Moving Consumer Goods
FoL	Formalization of Learning
GDPR	General Data Protection Regulation

IDMP	Identification of Medicinal Products
i.e.	that is (Latin: id est)
IoT	Internet of Things
IP	Information Production Process
IS	Information Systems
ISO	International Organization for Standardization
ISST	Institute for Software and Systems Engineering
IT	Information Technology
KPI	Key Performance Indicator
MES	Manufacturing Execution System
MDM	Master Data Management
MIT	Massachusetts Institute of Technology
ML	Machine Learning
MSP	Multi-sided Platform
MVOT	Multiple Versions Of Truth
OMG	Object Management Group
PF	Problem Formulation
PMI	Philip Morris International Inc.
PMO	Project Management Office
RaL	Reflection and Learning
RBV	Resource-based view
SEI	Software Engineering Institute
SMAC	Social, Mobile, Analytics, Cloud
SSOT	Single Source Of Truth
TDQM	Total Data Quality Management
TIQM	Total Information Quality Management
TQM	Total Quality Management
UML	Unified Modeling Language

1. Introduction

This doctoral dissertation describes the design and use of a capability reference model for strategic data management. It specifically addresses the fact that research so far has not come up with a reference model allowing today's data-driven enterprises, being in a process of digital transformation, to manage data as a strategic resource. The reference model defines and specifies design areas of data management in order to support users in the design, assessment, and improvement of specific data management activities. This introductory chapter gives an overview of the dissertation by outlining the motivation that stood behind the research conducted (*Section 1.1*), specifying the research questions, the research results, and the target audience of the dissertation (*Section 1.2*), and presenting the structure of the dissertation (*Section 1.3*).

1.1 Motivation and Problem Description

The need to establish digital business models, allowing companies to pursue data-driven business strategies and fact-based decision-making, is transforming today's enterprises, and even entire industries. Companies increasingly regard data as a key resource enabling new products and services (Dhar, Jarke, & Laartz, 2014; Lycett, 2013; Matt, Hess, & Benlian, 2015). Industry 4.0 scenarios in the manufacturing industry, for instance, are fundamentally transforming logistics and production processes; in the future, a smart product will steer itself self-dependently through the production process, order transportation services for being conveyed to the next workstation, and inform assembly units about what to do next (Hofmann & Rüsch, 2017; Kagermann, Wahlster, & Helbig, 2013). For this vision to become reality, data from multiple sources, including sensors, must be gathered, aggregated, enriched, provided, and maintained (Hermann, Pentek, & Otto, 2016). At the same time, companies collect and process an ever-increasing amount of customer data from external sources (such as corporate websites, social networks, or smart products) in order to provide an individualized customer experience and offer completely new, data-driven products and services (Lemon & Verhoef, 2016). A new term in this context is "datatization", denoting the transformation of an organization's capabilities and processes to change its value proposition by utilizing data analytics (Schüritz, Seebacher, Satzger, & Schwarz, 2017).

As the business relevance of data grows, companies are putting their focus more and more on the business value and business impact of data (Chen, Chiang, & Storey, 2012; Clarke, 2016). At the board level, this development is reflected by the establishment of a new role: the Chief Data Officer (CDO), who promotes the importance and relevance of data management across the entire organization, drives data-related topics, and is responsible and accountable for these topics at the executive level (Griffin, 2008; Lee et al., 2014; Xu, Zhan, Huang, Luo, & Xu, 2016). At the operational level, the data scientist and the data analyst (DalleMule & Davenport, 2017; Davenport & Patil, 2012) are two new roles that complement the work of data management and business intelligence (BI).

As industry increasingly turns its attention to data, governments and regulatory authorities are attaching more and more importance to the legal aspects of the rapidly evolving data economy. A major regulation is GDPR, the European Union's (EU) General Data Protection Regulation, which became effective in May 2018. GDPR strengthens privacy and protection of personal data, while at the same time imposing significant fines for non-compliant behavior of data collectors (cf. EU, 1995; EU, 2016). Compliance with regulatory requirements regarding data privacy and data security is thus becoming an even more relevant concern for companies collecting and using data in their day-to-day business activities.

Considering data a strategic resource represents a major challenge for enterprises, as they are required to question existing practices and rethink the way they manage data. A common understanding has been that data management comprises "policies, practices and projects that acquire, control, protect, deliver and enhance the value of data and information" (DAMA, 2009). In the 1990s, data management was a widely discussed topic both among practitioners and in the information systems (IS) research community. At that time, enterprise-wide systems (such as ERP systems) began to proliferate, allowing companies to break down data silos and interlink data on an organization-wide level (Davenport, 1998). Quality-oriented data management emerged as a new concept, which is still the basis of many reference models for data management used by companies today, as the quality of data has a significant impact on both business processes and decision-making processes (Batini, Cappiello, Francalanci, & Maurino, 2009; Madnick, Wang, Lee, & Zhu, 2009). By focusing only on the data quality dimension, however, these reference models largely neglect the strategic importance of data in terms of having value *for* business and having an impact *on* business. Despite the fact that data managers today regard aspects such as compliance, data security, agility, or big data as the biggest challenge of data management (Legner, Pentek, Ofner, & Labadie, 2017), the reference models in place do not address the business requirements of the data-driven enterprise.

The following example underlines the need for a data management reference model taking into account the data quality dimension *and* the strategic importance of data, i.e. its value *for* business and its impact *on* business:

> *Philip Morris International (PMI), a leading international tobacco company, is currently in a process of transforming its traditional business model, which has been to sell cigarettes via wholesalers to the retail industry (i.e. business-to-business, B2B), towards selling smoke-free products (e.g. electronic cigarettes) directly to the consumer (i.e. business-to-consumer, B2C). Data plays a prominent role in this strategic change of PMI's business model and product portfolio, since the new business model allows gathering valuable data about the consumers' requirements and preferences. This data can be used for continuous improvement of the products and permanent adjustment of the overall product and service offering. To facilitate this data-driven transformation, PMI decided to establish a central entity responsible for all data-related activities, covering both "traditional" aspects of data management and new forms, such as data science. A project team was established to review existing reference models for data management, with the aim to identify the model most suitable for PMI to assess the current situation and design a future-proof data organization. Since the project team did not succeed in finding a suitable reference model, however, it had to develop its own model.*

Another important aspect is how the importance of data management can be communicated throughout an organization, which is illustrated by the following example:

> *In 2016, Schaeffler, a global automotive and industrial supplier, received the CDQ Good Practice Award[1] for excellence in master data management (MDM). With a highly mature Corporate Data Management (CDM) organization already in place, the group meanwhile has extended data management from master data to other data domains. In view of Industry 4.0 being one of the key concerns of Schaeffler, the CDM team realized that it was crucial to establish some form of collaboration with Schaeffler's factory managers and also with Schaeffler's Digital Coordination Team orchestrating all activities related with digitization, such as Industry 4.0, for example. The three parties, not pursuing any considerable interaction with each other up to that point,*

[1] Launched by the Competence Center Corporate Data Quality (CC CDQ) and the European Foundation for Quality Management (EFQM) in 2013, a jury of international data management experts awards an innovative data management initiative every year. Further information can be found on www.cc-cdq.ch/cdq-good-practice-award.

> *agreed to conduct a workshop aiming at analyzing how CDM could support Industry 4.0 activities, particularly with regard to managing machine sensor data. During preparation of this workshop, the CDM team was searching for a reference model that could be used for describing the key elements of data management to non-experts.*

As these examples show, the data management community – both in industry and in research – has not been able to keep up with the emergence of new requirements posed by companies' digital transformation. Although a number of authors have dealt with aspects of data management going beyond data quality as the only, or predominant, dimension – such as data management for big data (Atul, Desale Girish, & Patil Swati, 2017; Chen et al., 2012; Chen et al., 2013), new roles required for data management (Davenport & Patil, 2012; Griffin, 2008; Lee et al., 2014; Xu et al., 2016), new methods (Lycett, 2013; O'Leary, 2014), or new application architectures (Chen, Li, & Wang, 2015; Smolander, Rossi, & Pekkola, 2017) –, a comprehensive reference model taking into account the specific requirements of data-driven enterprises is still missing. This gap in research, together with the growing business relevance of data and the increasing demand on the part of enterprises for conceptual and methodological support in establishing, assessing, and advancing data management, is what has motivated the studies presented in this dissertation.

1.2 Research Questions, Research Results, and Target Audience

Based on the understanding of data being a strategic resource of organizations, and data management being a dynamic capability, the dissertation describes the development and use of a capability reference model that allows enterprises to manage data as a strategic resource. This central topic of the dissertation can be broken down into two research questions:

- Research Question #1: What are the main design areas of strategic data management, and how are they interrelated?
- Research Question #2: Which concepts, models, and methods are suitable to support the establishment of strategic data management?

The dissertation addresses these research questions by following a design-oriented research approach. The methodology guiding the studies presented in this dissertation is design science research (DSR), which aims at solving practical problems while at the same time contributing to the advancement of the body of scientific knowledge. The results of applying this research method are called artifacts (Hevner, March, Park, & Ram, 2004). The dissertation describes the design, evaluation, and practical application of one artifact being the key outcome of the studies conducted: a capability reference model for strategic data management, named Data Excellence Model (DXM).

The DXM supports data managers in (a) defining a basic terminology of data management, (b) defining and specifying the main design areas and their respective deliverables, and (c) sharing knowledge regarding supporting concepts, models, methods, and good practices (Labadie & Legner, 2017). As a reference model, the DXM can be classified as “prescriptive knowledge” as described by Gregor’s (2006) “Type V” theory. It specifies the critical elements of a system – here: the main design areas of strategic data management – and offers orientation when it comes to designing a company-specific model (Fettke & Loos, 2007; Vom Brocke, 2007). In doing so, the capability reference model is the answer to both Research Question #1 and Research Question #2.

The focus of the research activities conducted for this dissertation was on global enterprises, which are typically characterized by geographically distributed operations and complex organizational structures (Roche, 1996). In these enterprises, the challenges of establishing and coordinating enterprise-wide strategic data management are particularly salient.

In accordance with DSR, the dissertation contributes both to the advancement of the body of scientific knowledge and to the broadening of the knowledge base of the practitioners’ community. Consequently, the target audience of the dissertation are both researchers and practitioners:

- The **research community** will benefit from this dissertation since its results advance the body of scientific knowledge in the field of data management. The reference model lays the foundation for further research by explicating the design areas of strategic data management and promoting the scientific debate to extend the quality-oriented perspective of data management towards a strategic view. From a conceptual perspective, the reference model adds the data management perspective to the academic discussion about organizational capabilities and the resource-based view (RBV). As far as methodology is concerned, the dissertation demonstrates how action design research (ADR), which is a combination of DSR and action research (AR), can be applied in a longitudinal, multilateral setting as the DXM is the culmination of numerous research results produced in the course of the research activities of the Competence Center Corporate Data Quality[2] (CC CDQ) since 2006.

- **Practitioners** in the data management profession (e.g. CDOs or data managers[3]) will benefit from the research results presented in this dissertation as they will be able to reflect their own approach of data management against the reference model presented, and by applying the reference model as a "blueprint" for establishing strategic data management in their enterprise. In addition, practitioners can use the reference model as a tool for communicating and explaining the elements of data management both to other data experts and to non-experts within the enterprise. Further, the DXM provides the foundation for a maturity model[4], which allows data management professionals to assess the maturity of their data management organization and to identify potentials of improvement.

1.3 Structure of the Dissertation

The dissertation is organized into eight chapters, constituting three main parts (see *Figure 1-1*):

(1) research context (*Chapters 1-3*),

(2) research activities and results (*Chapters 4-7*), and

(3) conclusion and outlook (*Chapter 8*).

Following this introduction, *Chapter 2* defines the theoretical and conceptual foundations of the dissertation. For this purpose, the chapter outlines the following concepts: digital transformation and the data-driven enterprise; data and data management; the resource-based view and the capability concept; and reference models and maturity models.

Chapter 3 provides an overview of existing reference models used for data management, originating both from academia and practice. It outlines the criteria underlying the author's decision to include or exclude certain models into/from his analysis, describes

[2] The CC CDQ is a consortium research program in the field of data management (see *Subsection 4.2*). Industry partners are more than 15 renowned companies from different industries. The researchers involved come from the University of Lausanne and the University of St. Gallen. Further information can be found on www.cc-cdq.ch.

[3] The user groups of the reference model are further specified in *Section 6.4*.

[4] The development of a maturity model based on the DXM is presented in *Subsection 5.5.1* as these activities influenced the contents of the DXM design areas (i.e. the success criteria and recommended practices). However, this dissertation's focus is on the reference model as the key artifact of the studies conducted.

and assesses each model deemed more or less relevant for this dissertation briefly in a separate subsection, and presents a comparative overview of the models selected. At the end of the chapter, the author derives implications from this evaluation to guide the research conducted for this dissertation.

Chapter 4 presents the research design underlying the construction of the capability reference model as the key artifact of this dissertation. It presents the research methodology and techniques applied, introduces the CC CDQ, being the consortium research program which the research activities were embedded in, and then details the concrete research activities.

Chapter 5 presents the case studies conducted and evaluated by the author of the dissertation. It details the experiences and results gained from three exploratory case studies, which had a substantial impact on the design of the reference model, as key requirements were derived from these cases. Furthermore, it demonstrates the applicability and practical utility of the DXM with the help of seven case studies reflecting two typical scenarios for applying the reference model in practice: (1) translating the abstract design knowledge of the reference model into a concrete situational design, and (2) applying the reference model as abstract situational knowledge for communication, education, maturity assessment, and benchmarking purposes.

Chapter 6 outlines the process of designing the key artifact presented in this dissertation: the DXM as a capability reference model for strategic data management. It provides an overview of the design requirements for the artifact, introduces the reference model by detailing its nature, structure, and meta-model, and elaborates on the design decisions the author made during the research process. Furthermore, the chapter presents a description of the data management roles specifically addressed by the DXM, and how each of these roles can benefit from using the reference model. Finally, the author describes each design area of the DXM on a detailed level.

Chapter 7 presents an evaluation of the reference model by means of three distinct evaluation strategies: (1) naturalistic evaluation, (2) formal evaluation, and (3) comparison with competing artifacts.

Chapter 8 summarizes the results of this dissertation and its contribution both to the advancement of the body of scientific knowledge and to the broadening of the knowledge base of the practitioners' community. Finally, it points to the limitations of the research results and proposes actions for future research.

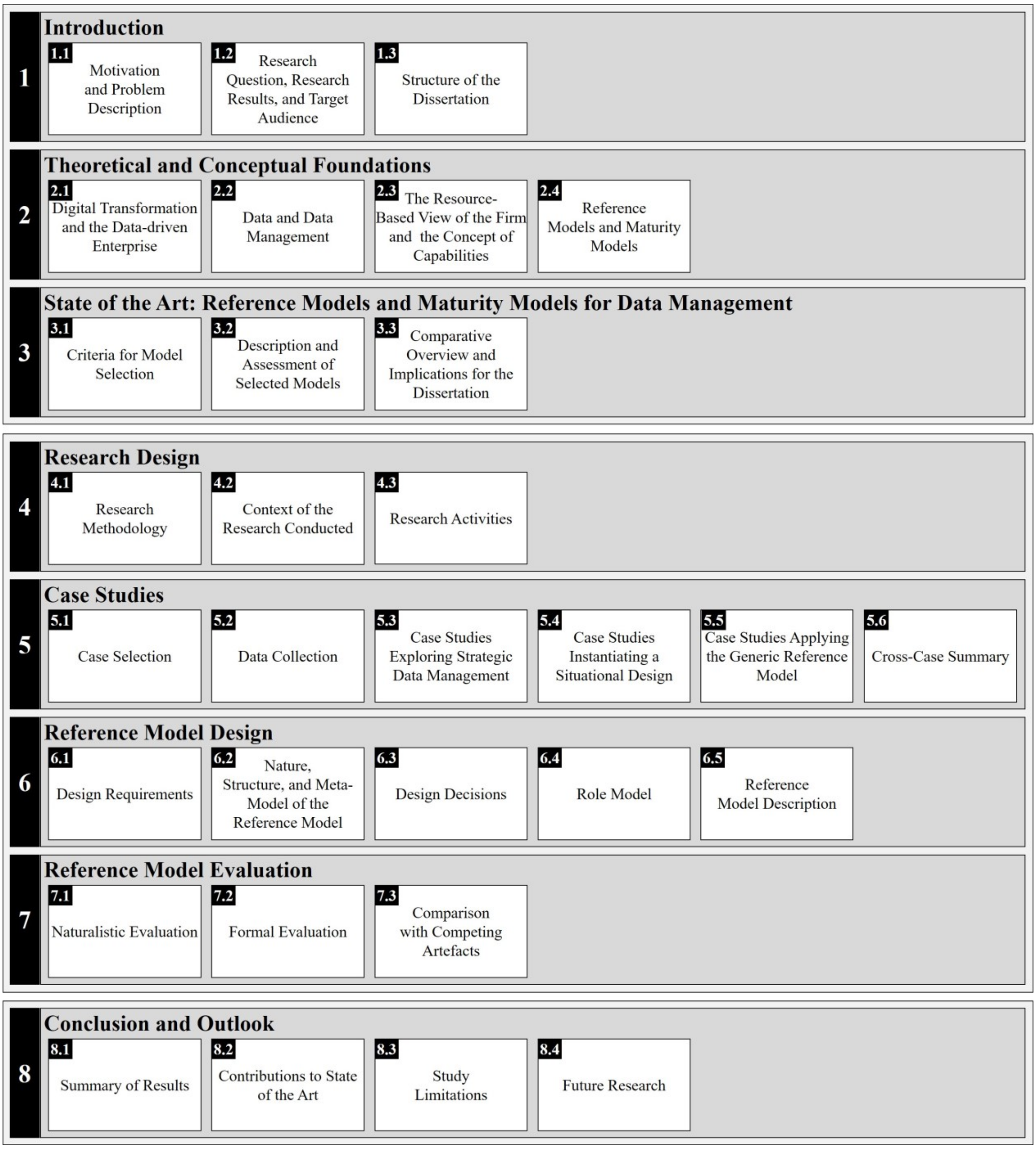


Figure 1-1: Structure of the dissertation

2. Theoretical and Conceptual Foundation

This chapter presents the scientific concepts, theories, and approaches building the foundation of the studies conducted for this dissertation. These come from both the IS domain and the general management domain. The four sections of this chapter reflect the central research question (i.e. how enterprises can manage data as a strategic resource) and the title of the dissertation (Capability Reference Model for Strategic Data Management) by describing and explicating the contextual setting motivating the research activities (*Section 2.1*), the subject of analysis (*Section 2.2*), the theoretical perspective from which the subject of analysis is examined (*Section 2.3*), and the artifact constituting the concrete result of the research conducted (*Section 2.4*).

2.1 Digital Transformation and the Data-Driven Enterprise

The terms "digital transformation" and "data-driven enterprise" describe two prevailing, complementary developments that can be observed in the corporate world: (1) the fundamental transformation of business models, business processes, products, and services caused by the use of advanced digital technologies (Leimeister, Österle, & Alter, 2014; Matt et al., 2015); and (2) the practice on the part of enterprises to pursue data-driven strategies and fact-based decision-making with the help of big data and data analytics (Buhl, Röglinger, Moser, & Heidemann, 2013; Davenport, 2014; Dhar et al., 2014; Lycett, 2013; Provost & Fawcett, 2013). The following subsections detail the terms and concepts of *digital transformation* (*Subsection 2.1.1*) and the *data-driven enterprise* (*Subsection 2.1.2*).

2.1.1 Digital Transformation

The term "digital transformation" refers to the digitalization[5] of processes and business models in the corporate world, but also to the digitalization of our everyday life (Matt et al., 2015). Digital transformation is mainly driven by technological progress, leading to falling prices for sensors and facilitating cheaper and faster collection, processing, and sharing of data (Fleisch, Christ, & Dierkes, 2005). This development resulted in the convergence of the so-called SMAC technologies (i.e. social, mobile, analytics, and cloud technology) (Bardhan, Demirkan, Kannan, Kauffman, & Sougstad, 2010; Leimeister et al., 2014). Digital transformation is a "megatrend" including various topics,

[5] The terms *digitalization* and *digitization* are often used interchangeably both in practice and in the scientific discourse (Bärenfänger (2017). This dissertation uses the term *digitalization*.

such as *digital platforms*, the *Internet of Things (IoT)*, or *Industry 4.0*[6]. To illustrate what digital transformation is about, these three topics are briefly outlined in the following paragraphs.

Digital Platforms

Digital transformation of business models, business processes, products, and services has brought about many types of digital platforms, also called multi-sided platforms (MSPs) (Tan, Pan, Lu, & Huang, 2015). MSPs are defined as "technologies, products or services that create value primarily by enabling direct interactions between two or more customer or participant groups" (Hagiu, 2014, p. 4). Prominent examples are eBay (bringing together buyers and sellers), Airbnb (bringing together apartment owners and renters), Uber (bringing together people looking for a taxi service and private drivers offering that service), Facebook (interlinking users with each other but also with advertisers, game providers, and content developers), or the app stores of Google and Apple (relating application developers with potential application users). With the number of participants on such a platform growing, the benefit for each participant grows, as the costs for search and transaction decrease (Hagiu, 2014; Weill & Woerner, 2015).

Internet of Things (IoT)

The fact that digital transformation not only connects individuals and organizations on a global level, but also "things" (Bharadwaj, El Sawy, Pavlou, & Venkatraman, 2013), led to the emergence of the term IoT. Based on technological progress and decreasing costs for microprocessors, things (i.e. machines, devices, or products) receive a digital identity by which they can be identified and addressed unambiguously (Fleisch et al., 2005). In combination with sensors, the IoT generates "a steady stream of information about where devices are, how they're being used, their condition, and the state of their environment" (Williams, 2014, p. 1). This allows enterprises to offer "smart" products, interact with consumers over the entire product lifecycle, and provide them targeted, personalized after-sales services based on how the consumer has used the product.

Industry 4.0

The term "Industry 4.0" stands for the convergence of industrial production and information and communication technology. Since the German federal government announced "Industrie 4.0" as one of the key initiatives of its high-tech strategy in 2011

[6] According to Bärenfänger and Otto (2015), further digitalization aspects are mobile computing, big data, and hybrid products.

(Kagermann et al., 2013), Industry 4.0 has been, and still is, one of the most frequently discussed topics among practitioners and academics not only in Germany, but also in the German-speaking area and beyond[7]. Numerous articles, both by researchers and practitioners, and conferences have dealt with that topic (Vogel-Heuser, Bauernhansl, & Hompel, 2017). Generally speaking, Industry 4.0 aims at the efficient industrial production of customizable products (i.e. mass customization) (Bücker, Hermann, Pentek, & Otto, 2016). Enabled by the communication between people, machines, devices, and sensors, Industry 4.0 means a paradigm shift from centrally controlled production processes of mass products to decentralized production of customizable products. Smart, customized products have their own production history electronically stored on board, as well as information about their current state and target state, and actively steer themselves through the production process by instructing machines to carry out the next step in the manufacturing process and ordering transportation services for being conveyed to the next production unit (Kagermann, 2015). Hermann et al. (2016) have introduced four design principles of Industry 4.0 scenarios:

(1) **interconnection**, understood as the ability of machines, devices, sensors, and humans to connect and communicate with each other (e.g. via the IoT);

(2) **information transparency**, meaning the ability of information systems to create a virtual copy of the physical world by enriching digital plant models or product models with sensor data;

(3) **decentralized decision-making**, referring to the ability of smart products and machines to make decisions on their own and perform tasks as autonomously as possible;

(4) **technical assistance**, referring to (1) the ability of information systems to support humans by providing aggregated and visualized information, allowing them to make well-informed decisions and solve urgent problems at short notice; and (2) the ability of smart machines to physically support humans by taking over a range of tasks that are either unpleasant, too exhausting, or too dangerous for being carried out by humans (pp. 3932-3933).

In the context of Industry 4.0 scenarios, data sovereignty plays a key role when it comes to companies sharing and exchanging data with multiple partners. Data sovereignty – defined as a "corporate entity's capability of being entirely self-determined with regard

[7] Related concepts to Industry 4.0 are "Industrial Internet", "Advanced Manufacturing", "Integrated Industry", "Smart Industry", and "Smart Manufacturing" (Hermann et al. (2016).

to its data" (International Data Spaces Association, 2019, p. 9) – empowers companies to keep control over their data assets by defining data usage policies other companies must obey when using their data. This protects data ownership rights, while at the same time fostering the willingness of companies to share and exchange data among each other.

As these short illustrations with regard to digital platforms, the IoT, and Industry 4.0 suggest, the digital transformation is a multi-faceted phenomenon. At its core, enterprises leverage state-of-the-art digital technology to improve their performance and extend their offering through the development of digital business models, business processes, products, and services. The importance of the digital transformation megatrend is reflected by a new role being established in enterprises: the Chief Digital Officer[8], who is responsible and accountable for digital transformation of the enterprise at the board level and closely interacts with its IT counterparts (Horlacher & Hess, 2016).

2.1.2 The Data-Driven Enterprise

In view of the ever-increasing amount of data as a result of digital transformation, enterprises are striving for new ways to gain value from all that data. They use powerful technologies – such as big data and data analytics – in order to pursue data-driven strategies and fact-based decision making (Provost & Fawcett, 2013). While data keeps originating from internal enterprise systems, it increasingly comes from web and IoT sources as well (Porter & Heppelmann, 2015). For instance, companies collect large volumes of customer data as click streams and feeds from corporate websites and social networks, but also from sensors in production machines and "smart" products. They then analyze this data in order to be able to provide a tailored customer experience and offer completely new, data-driven products and services (Lemon & Verhoef, 2016). The transformation of the capabilities of an organization and the change of its value proposition through data analytics has been denoted by the terms "datatization" (Schüritz et al., 2017) and "data monetization" (Wixom & Ross, 2017). With datatization/data monetization, enterprises benefit from big data and data analytics to innovate their business in various ways, such as (1) data-enabled process improvements, aiming to optimize processes, increase productivity, support decision-making, and gain insights about customers; (2) data-enriched products and services; and (3) data-driven services, providing new value propositions for enterprises (Schüritz et al., 2017; Wixom & Ross, 2017).

---

[8] In parallel to the *Chief Data Officer*, the role of the *Chief Digital Officer* is also abbreviated with *CDO*. This dissertation denominates the *Chief Data Officer* as *CDO*.

As a consequence of the increasing strategic importance of data, enterprises are more and more affected by regulations (Stockdale, 2014). A major regulatory provision is the EU's General Data Protection Regulation (2016), which became effective in May 2018. GDPR significantly strengthens the protection of personal data and imposes significant fines for non-compliant behavior. It basically aims at "giving the power back" to data subjects (e.g. customers revealing personal data when filling out an electronic form on a website) by providing them the opportunity to define precisely what data collectors may do with their data and what not[9]. Initiated by the EU as a framework for harmonizing data privacy policies across all EU member states, GDPR requires companies to seek for explicit and unambiguous consent from data subjects and implement the "right to be forgotten". Moreover, data collectors must maintain adequate security mechanisms for all critical data in order to avoid data misuse, data leaks, and unauthorized data access. But not only regulators and lawmakers are having an eye on data privacy and data security, also data subjects – such as customers directly providing personal data and indirectly providing transactional data when they buy online, for example – and the general public have become increasingly sensitive about how personal data is used by data collectors (Zwitter, 2014). Handling data in a compliant way also requires enterprises to ensure data security by protecting not only their technical infrastructure but also data, for which access rights must be defined (O'Brien, 2014).

The development towards the data-driven enterprise is reflected on the organizational level by the establishment of new roles, such as the CDO, the data scientist, and the data analyst (DalleMule & Davenport, 2017; Davenport & Patil, 2012; Griffin, 2008; Lee et al., 2014; Xu et al., 2016). On the process and technical level, enterprises increasingly need to be able to handle *big data* and apply *data analytics* (Provost & Fawcett, 2013). These two concepts are briefly outlined in the following paragraphs.

Big Data

The term "big data" refers to the observation that both digital transformation and data-driven innovation require and generate large amounts of different types of data coming from different sources and being transferred at a rapid speed. A frequently used pattern for describing the concept of big data is the "3 Vs" (i.e. volume, velocity, and variety)[10] (Bärenfänger & Otto, 2015; Chen et al., 2012; Waller & Fawcett, 2013).

[9] In that way, GDPR imposes data sovereignty for person-related data.

[10] *Veracity* has been added as the fourth "V" of big data (Abbasi, Sarker, and Chiang, 2016; Buhl et al., 2013), referring to the challenge of gaining trustworthy results from big data, knowing that data quality is often low.

- *Volume* refers to the increasing amount of data available from various sources[11].
- *Velocity* describes the higher frequency for gathering data from the sources (e.g. via data streams) and short periods of time required for processing data (e.g. in (near) real time).
- *Variety* hints at the various sources of data and the multiple formats and structures, in which data is available.

Data Analytics

Coping with the increasing volume, velocity, and variety of data requires specific technical infrastructures, such as enterprise analytics platforms (Abbasi et al., 2016; Chen et al., 2015; O'Leary, 2014), and organizational capabilities. Kwon, Lee, and Shin (2014) define data analytics as an "innovative IT capability that can improve firm performance" (p. 387). Data analytics aims at provision of data-based insights for justifying, guiding, and prescribing business activities by capturing, aggregating, analyzing, and disseminating data (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). The goal is to maintain an enterprise's competitiveness and facilitate operational decision-making (Davis, 2014).

In order to use "big data" in data analytics activities, two technical solutions – in addition to traditional data warehouse databases and business intelligence (BI) applications – have gained importance in data-driven companies: (1) enterprise analytics platforms and (2) data catalogs. In merely descriptive data analytics scenarios, in which the intended use of data is known upfront, data is cleansed and loaded into a data warehouse or BI application in a predefined schema in order to provide a single source of truth (SSOT).

(1) In more advanced data analytics scenarios, in which the exploration of data generates previously unknown insights and recommendations, data is loaded into an enterprise analytics platform in an undefined structure to provide multiple versions of truth (MVOT) for multiple (future) data usage scenarios. Enterprise analytics platforms allow users to explore data, develop, and run data pipelines to generate different data (analytics) products, such as reports and ad-hoc analysis for descriptive data analytics scenarios, or data science products for predictive and prescriptive data analytics scenarios (Fadler & Legner, 2019).

[11] See *Subsection 2.2.2* for an overview of these data sources.

(2) Data catalogs create transparency regarding the data that is available within a company, and provide access to this data. As curated platforms, data catalogs match data supply (i.e. cataloged data) and data demand (i.e. data discovery and access) (Korte, Fadler, Spiekermann, Legner, & Otto, 2018).

While the technology allowing efficient processing of big data is new, the mathematical and statistical methods required for doing so are long since known (Chen et al., 2012; Davis, 2014). Artificial intelligence (AI) and machine learning (ML), for example, are methods that have been developed over decades, but could not be extensively used due to technological restrictions (Langley & Simon, 1995). Besides being a result of the technological progress made in recent years (Fleisch et al., 2005), the growing adoption of data analytics is also due to the fact that it has become more and more accepted on all organizational levels in enterprises (LaValle et al., 2011).

In essence, the *data-driven enterprise* transforms its business model, products, services, and processes through digital technologies and applies fact-based decision-making. Both digital transformation and data-orientation significantly change the goals, priorities, and organization of enterprises (see *Table 2-1*).

Table 2-1: Characteristics of digital transformation and the data-driven enterprise

	Digital transformation	**Data-driven enterprise**
Objectives	- Use of digital technologies to radically improve business operations - Better performance and reach of the enterprise	- Use of BI and (big) data analytics techniques - Exploitation and exploration of data for business value creation
Results	- Digital business models, digital products and/or services - Digital consumer experience - Digital business operations	- Data-enabled process improvements and decision-making - Data-enriched products and services - Data-driven services
Responsible functions/ roles	- Chief Digital Officer - Digital initiatives	- Chief Data Officer - Data scientists/analysts

2.2 Data and Data Management

This section deals with the basic characteristics of data (*Subsection 2.2.1*), different data types (*Subsection 2.2.2*), the notion of data as being an economic good (*Subsection*

2.2.3), and the concept of data quality (*Subsection 2.2.4*). It also provides an overview of data management and its evolution from being quality-oriented towards becoming of strategic importance (*Subsection 2.2.5*).

2.2.1 Data

Generally speaking, data is used to describe real-world objects and their characteristics (Boisot & Canals, 2004). A real-world object (e.g. the customer of a company) is represented by a data object, the characteristics of which (e.g. the customer's name or address) are in turn represented by attributes (Mertens et al., 2017). With regard to the difference between "data" and "information", there has always been a vivid discourse both in the practitioners' and the scientific community. However, a generally accepted distinction of the two concepts still does not exist (Boisot & Canals, 2004). The dominant (information processing) view argues that data turns into information when it is processed or used in a certain context (Krcmar, 2015; March & Smith, 1995; Oppenheim, Stenson, & Wilson, 2003; van den Hoven, 1999). Hence, several contributions describe data as the raw material of information (cf. DAMA, 2009; Wang, 1998). Processing information and connecting it with other information finally results in knowledge (DAMA, 2009; Krcmar, 2015).

In practice, it is difficult (and in many cases not possible) to distinguish between data and information, as both can directly be accessed and processed in information systems (Hansen, Neumann, & Mendling, 2015). In line with other scientific publications on data management – such as Pipino, Wang, Kopcso, and Rybolt (2014) or Wang (1998) – this dissertation does not differentiate between "data" and "information", and only applies the term "data".

2.2.2 Types of Data

As for categorizing data into different types, the IS domain in the past was mainly focusing on company-internal data (Leimeister, 2015). "Traditional" typologies had data categorized into (1) *master data*, (2) *transactional data*, (3) *inventory data*, and (4) *change data*. These four types show differences in terms of data change frequency, time reference, and volume (Hansen et al., 2015; Kokemüller & Weisbecker, 2009; Leimeister, 2015):

(1) *Master data* describes the core business objects of an enterprise. Master data sets are relatively stable over time and can be referenced by transactional data (Dreibelbis, 2008; Loshin, 2009). Common examples of master data are

material master data, product master data, supplier master data, customer master data, and employee master data.

(2) *Transactional data* represents documents generated and used in business processes, such as invoices or purchase orders, which describes the changes when updating inventory data (Kokemüller & Weisbecker, 2009).

(3) *Inventory data* provides information on stock or account level. Inventory data sets are subject to frequent change (Hansen et al., 2015). Common examples are the number of products available on stock or bank account balances.

(4) *Change data* informs about changes made when updating master data[12] (Stahlknecht & Hasenkamp, 1999).

Alongside with these "classical" data types, two other types of data are important in data management:

(5) *Reference data* is often referred to as externally defined data (DAMA, 2017). Other authors state that there is no unambiguous definition for this type of data (Otto, 2012a). Examples of reference data are currency codes or country codes, as specified by the International Organization for Standardization (ISO).

(6) *Metadata* is data about data. Among other things, metadata defines structural and formal requirements for data entry and data governance, such as data ownership or data access rights (Hodge, 2004; Sen, 2004).

While the above introduced six data types mainly originate from company-internal sources, data originating from external sources has become more and more relevant for companies in the wake of digital transformation and the growing tendency on the part of enterprises to become data-driven. These are mainly (a) *open data*, (b) *usage data*, (c) *community data*, and (d) *sensor data*.

a. *Open data* is made available for free by governmental institutions or (non-profit) organizations (Sadiq & Indulska, 2017). Examples are weather or flight data.
b. *Usage data* results from the digital activities of users, such as browsing a website or using the smartphone for various purposes.
c. *Community data* is produced by users interacting on social media platforms.

---

[12] Similarly, transactional data describes the changes when updating inventory data.

d. *Sensor data* is produced by the use of sensors in smart products, such as fitness trackers, or smart machines used in Industry 4.0 settings (Baesens, Bapna, Marsden, Vanthienen, & Zhao, 2016; George, Haas, & Pentland, 2014).

2.2.3 Data as an Economic Good

How important data is for enterprises has been discussed since the 1980s (Goodhue, Quillard, & Rockart, 1988), when electronic data became ubiquitous in the business world. This led to the notion of data as being an economic good, viewed either as a (1) product, (2) an asset, or (3) a resource.

(1) Viewing data as a product draws an analogy between information products and physical products. In analogy with raw material being the starting point for the production of a physical good, raw data processed in information systems can be considered the starting point for producing an information product (Lee, Pipino, Funk, & Wang, 2006; Wang, 1998).

(2) Viewing data as an asset means considering data an intangible asset that has a period of useful life during which it is beneficial for an enterprise (Horne, 1995; Oppenheim et al., 2003).

(3) Viewing data as a resource basically means that data can be characterized by certain features typical of any resource, such as intangibility, divisibility, and transportability (Levitin & Redman, 1998).

2.2.4 Data Quality

With the development towards enterprise-wide process and data integration in the 1990s, academics and practitioners likewise began to recognize data quality as a critical aspect affecting the value of data. Various studies have revealed that data quality has an impact on business processes, such as in supply chain management (Tellkamp, Angerer, Fleisch, & Corsten, 2004; Vermeer, 2000), in customer relationship management (Reid & Catterall, 2005; Zahay & Griffin, 2003), regarding BI (Orr, 1998; Price & Shanks, 2005; Shankaranarayanan, Ziad, & Wang, 2003), or regarding the enterprise's overall performance (Redman, 1995; Redman, 1998; Sheng, 2003; Sheng & Mykytyn, 2002; Wamba, Akter, Trinchera, & Bourmont, 2018). Over the past decades, various scholars then specified the concept of data quality both from the database perspective and the management perspective. A widespread view emerging during this period of time states that the quality of a data object cannot be determined by means of a single criterion, but that data quality is a multi-faceted concept comprising various *data quality dimensions*, such as consistency, completeness, accuracy, or timeliness (Wang & Strong, 1996).

As all these dimensions are context-specific, a new approach began to emerge towards the end of the 1990s: data's "fitness for use" (Otto, 2011b; Redman, 2001; Wang & Strong, 1996). According to this view, whether data fits a certain purpose depends on the perception and expectation of the data user with regard to the specific context. For example, whereas for aircraft maintenance records or bank account statements a 100-percent data accuracy is crucial, an 80-percent accuracy might be sufficient for an employee's home phone number (Moody & Walsh, 1999).

2.2.5 Data Management

Data management can be defined as "the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data" (DAMA, 2017, p. 17). Its core purpose is to develop a strategy and define the organizational responsibilities for maintaining, capturing, and providing data in a coordinated way (Krcmar, 2015). According to Otto and Österle (2015), data management "makes decisions and executes measures that affect the company-wide handling of data" (p. 192). Furthermore, they state that "quality management of master data is among the most important sub-tasks of data management" (Otto & Österle, 2015, p. 192).

In line with the view of Otto and Österle (2015), data management mainly used to focus on data quality in the past. However, with the role of data changing in enterprises, there has been a growing tendency towards considering the strategic importance of data management. Both approaches are introduced in the following paragraphs.

Quality-Oriented Data Management

For a long time, data management focused on providing high-quality data (Kahn, Strong, & Wang, 2003; Otto, 2011b; Wang, 1998), with the following key contributions:

The first integrated approach regarding data quality was Total Data Quality Management (TDQM). It originated from a research program initiated by Richard Wang at the Massachusetts Institute of Technology (MIT). Based on the notion of data being viewed as a product, TDQM transfers elements from Total Quality Management (TQM)[13] to

[13] TQM describes a quality-centered management approach aiming at customer satisfaction through providing products and services of highest quality by involving all members of the organization in quality-related activities.

information production processes (IP). This so-called “IP approach” proposes four principles: (1) understand consumers’ information needs, (2) manage information production processes, (3) manage information lifecycles, and (4) appoint information product managers responsible for these activities (Wang, 1998; Wang, Lee, Pipino, & Strong, 1998).

The next approach was Total Information Quality Management (TIQM), addressing mainly data warehouse projects by adopting also a “data as a product” view. TIQM seeks to meet all demands of data consumers by creating transparency regarding quality requirements along the “data supply chain” (English, 2003).

Another prominent concept was the Complete Data Quality Methodology (Batini & Scannapieca, 2006), considering data quality management as the sum of all activities of data quality improvement in an enterprise, and seeking to improve business processes under consideration of an optimal cost-benefit ratio.

As a response to the previously mentioned concepts, researchers from the University of St. Gallen in 2006 established the CC CDQ. The Competence Center adopts an organizational perspective, which is largely neglected by the preceding approaches, which the St. Gallen researchers criticized also for providing methodical support for certain data management activities only (Otto, Wende, Schmidt, & Osl, 2007; Weber, Otto, & Österle, 2009b). The CC CDQ’s reference model for Corporate Data Quality Management (CDQM) understands quality-oriented data management as a dynamic organizational capability comprising six design areas: strategy, controlling, organization, processes, architecture, and applications (Otto et al., 2007; Otto, 2011b; Otto & Österle, 2015; Schemm, 2008).

In a nutshell, quality-oriented data management can be summed up as follows (see *Table* 2-2 at the end of this section)[14]:

- Its goal is to improve data quality by applying the *philosophy and methods from quality management*. As a comprehensive management approach, it introduces a quality management system for data encompassing strategic, organizational, and technical aspects.
- Quality-oriented data management approaches introduce a *product view on data, along with the related information production processes*. They focus on storing and distributing data in the enterprise, emphasizing data modeling, data

[14] For a review of data quality research, see Batini et al. (2009) and Madnick et al. (2009).

architecture (Goodhue et al., 1988), and applications (Ballou, Wang, Pazer, & Tayi, 1998).

– The asset or resource view in data management emphasizes the organizational prerequisites to improve and maintain data quality. It highlights that improving data quality requires *data governance*, which describes the assignment of decision rights and responsibilities (Khatri & Brown, 2010; Weber et al., 2009b). Key responsibilities for quality-oriented data management are mainly attributed to dedicated data management or BI teams (Panian, 2010). The responsibilities are explicated in the form of data management processes (Kahn et al., 2003), which include data quality controlling processes (Pipino, Lee, & Wang, 2002; Wang, 1998).

Although there is general consensus that poor data quality negatively impacts business performance, enterprises still struggle with raising the level of data quality. Based on a review of empirical studies, Marsh (2005) summarizes that "88 per cent of all data integration projects either fail completely or significantly over-run their budgets, […] 33 per cent of organisations [sic] have delayed or cancelled new IT systems because of poor data, […] Less than 50 per cent of companies claim to be very confident in the quality of their data" (p. 106). To address the challenges organizations face in improving data quality, several data management frameworks, reference models, and maturity models have been developed by researchers and practitioners (see *Chapter 3*).

Strategic Data Management

Since the 2010s, the perspective on data in academia and in practice has been impacted by the megatrend of digitalization and the transformation of businesses into data-driven enterprises – as outlined in *Section 2.1*. Both developments promote the view of data as a strategic resource, and complement the prevailing perception of data in several ways (see *Table 2-2* at the end of this section):

- They emphasize data's *business value and impact* (Chen et al., 2012; Clarke, 2016). Organizations benefit in various scenarios from data and analytics to innovate their business (Schüritz et al., 2017; Wixom & Ross, 2017): (1) data-enabled improvements, which seek to optimize processes, increase productivity, support decision-making, and create insights about customers, (2) data-enriched products and services, and (3) data-driven services, which both provide new value propositions for enterprises.

- These scenarios require and generate *data of large volume, high velocity, and high variety* (Bärenfänger & Otto, 2015; Chen et al., 2012; Waller & Fawcett, 2013). These data originate not only from internal and proprietary enterprise systems, but are increasingly unstructured data (i.e. text, video, audio) and data from web and IoT sources (Porter & Heppelmann, 2015). Acquiring, storing, and processing the increasing variety and volume of data requires additional infrastructures, such as enterprise analytics platforms (Abbasi et al., 2016; Chen et al., 2015; O'Leary, 2014) and data catalogs (Korte et al., 2018).
- DalleMule and Davenport (2017) highlight the need for a data strategy for any company, including both defensive and offensive aspects of data management. While defensive elements aim at establishing control over data (by ensuring data security, privacy, integrity, quality, regulatory compliance, and governance), offensive elements address the way data is used in order to gain competitive edge and increase profitability.
- On the organizational level, data's strategic role is reflected by a new role: the CDO, who introduces organizationally sanctioned leadership and accountability for data-related topics at the executive level, compared to the lower-level data managers (Griffin, 2008; Lee et al., 2014; Xu et al., 2016). In addition, enterprises have to build new skill sets and competencies in the domain of big data and advanced analytics, with the roles of data engineers and data scientists (Davenport & Patil, 2012).
- In addition to the enterprise-internal attention on the data resource, governments and regulatory authorities are attaching increasing importance to the legal issues of the data economy. Among the emerging regulations are the European GDPR, which strengthens and unifies protection of personal data and imposes significant fines for non-compliant behaviors (1995; 2016). Thus, compliance with regulatory requirements as well as considerations of data privacy, sovereignty, and security are becoming relevant concerns for data management, in addition to data quality.

To address the changing role of data, researchers have highlighted that establishing advanced analytics requires building individual, organizational, and technical capabilities (Schüritz et al., 2017) for collecting, storing, analyzing, and communicating big data from multiple external and internal data sources (Chen et al., 2012). By reviewing the big data value chain, Abbasi et al. (2016) outline a research agenda for deriving knowledge, decisions, and actions. Specifically, they encourage researchers to address privacy and security concerns, organizational culture and governance, big data outcomes, and business process improvements and automation. Some attempts have been made by researchers to provide a comprehensive perspective on big data capabilities and resources, such as the Big Data Analytics Capability Model (Akter, Wamba, Gunasekaran, Dubey, & Childe, 2016) and the Big Data Resources Framework (Gupta & George, 2016) (see *Chapter 3*). However, these endeavors remain isolated and do not build on the large body of data management-related knowledge that has been built in the academic and practitioner communities over the past two decades.

Development and Definition of Data Management

Table 2-2 provides an overview of the key characteristics of quality-oriented and strategic data management.

Table 2-2: Evolution of data management approaches

	Quality-oriented data management (since the 1990s)	**Strategic data management (since the 2010s)**
Responsible functions/roles	Master Data Management (MDM), Business Intelligence (BI)	Chief Data Officer (CDO)
Business context	Business process redesign and integration	Digitalization/digital transformation, data-driven innovation/data economy
Role of data	Data as an enabler of a. Operational excellence (i.e. business processes and decision-making)	Data as an enabler of a. Operational excellence (i.e. business processes and decision-making) b. New/enhanced business models (i.e. data-enriched products/services and data-driven services) c. Risk reduction and compliance
Management objective	a. Provision of high-quality data	b. Business value generation through data
Data-related concerns	a. Data quality	a. Data quality b. Compliance c. Data privacy d. Data security
Relevant data sources	a. Internal sources: (mainly proprietary) such as enterprise systems (ERP, BI)	a. Internal sources b. External sources: open data, usage data, community data, sensor data

Based on the above described perspectives on data management and in line with Legner, Pentek, and Otto (2020), this dissertation defines data management as a socio-technological design task comprising strategic, organizational, and technological aspects to manage data in an enterprise. Reflecting the changing role of data from an enabling to a strategic resource, strategic data management aims at generating business value while considering compliance, data privacy, data security, and data quality of both internal and external data sources.

2.3 The Resource-Based View of the Firm and the Concept of Organizational Capabilities

The resource-based view of the firm (RBV) and its extension, the concept of organizational capabilities, provide the theoretical lenses through which the studies conducted for this dissertation are viewed. This section outlines both concepts and puts them in relation with data management.

2.3.1 Resource-Based View of the Firm

The RBV is a theory in strategy management (Henderson & Mitchell, 1997) claiming that enterprises can gain a competitive advantage by possessing and/or controlling certain resources (Barney, 1991; Spanos & Lioukas, 2001). According to Helfat and Peteraf (2003), a resource is an "asset or input to production (tangible or intangible) that an organization owns, controls, or has access to on a semi-permanent basis" (p. 999). Resources are either tangible (i.e. physical goods, such as production facilities or raw materials) or intangible (i.e. non-physical goods, such as patents, brand names, or usage rights). These resources have the potential to lead to a competitive advantage for the enterprise if they are valuable, rare, inimitable, and non-substitutable (Barney, 1991; Dierickx & Cool, 1989; Mata, Fuerst, & Barney, 1995).

2.3.2 Organizational Capabilities

The concept of organizational capabilities is an extension of the RBV, arguing that enterprises should not just possess and/or control certain resources but should also be capable of combining, developing, and/or utilizing these resources in order to gain competitive advantage (Bharadwaj, 2000; Collis, 1994; Wu, Melnyk, & Flynn, 2010). Thus, capabilities are value-adding activities making use of corporate resources. Wade and Hulland (2004) define capabilities as "repeatable patterns of actions [… which …] transform inputs into outputs of greater worth" (p. 109). Capabilities are contingent on business objectives as they are formed by the strategic goals of the enterprise (Amit & Schoemaker, 1993), and results-oriented as they aim at a particular outcome (Bharadwaj et al., 2013; Helfat & Peteraf, 2003).

Types of Organizational Capabilities

Capabilities can be distinguished by the degree to which they contribute value to an enterprise's competitiveness. The concept of organizational capabilities reflects this understanding and suggests a strategic use of capabilities (Moingeon, Ramanantsoa, Métais, & Orton, 1998), which, according to their value contribution and strategic importance, can be categorized as follows (Baghi, 2016):

- *Ordinary capabilities*[15] focus on specific business processes and/or functions, allowing enterprises to improve their overall performance.
- *Core capabilities*[16] allow enterprises to combine several ordinary capabilities and resources (Stoel & Muhanna, 2009). They are of strategic importance and help enterprises gain competitive advantage.
- *Dynamic capabilities*[17] allow enterprises to maintain or extend a competitive advantage in a changing market environment by developing, combining, or reconfiguring core capabilities, ordinary capabilities, and resources (Teece, Pisano, & Shuen, 1997; Wang & Ahmed, 2007). Examples of a dynamic capability are new product development or strategic decision-making (Eisenhardt & Martin, 2000).

Capability Lifecycle

Enterprises establish organizational capabilities in the form of *routines*, which implement certain tasks and activities within business processes (Marino, 1996; Setia, Setia, Venkatesh, & Joglekar, 2013). As organizations grow, their capabilities evolve over time. To address the dynamic development of capabilities, Helfat and Peteraf (2003) introduce the concept of the capability lifecycle. This concept considers the maturity of capabilities and proposes three evolution stages in the initial lifecycle of organizational capabilities:

---

[15] Other authors use the term *operational capability* (Helfat & Peteraf, 2003; Pavlou & El Sawy, 2011).

[16] Other authors use the terms *competencies* or *core competencies* (cf. Thomas, Heene, & Sanchez, 1996).

[17] Other authors use the terms *combinative capability* (Kogut & Zander, 1992), *first-order dynamic capability* (Pavlou & El Sawy, 2011), or *architectural competence* (Henderson & Cockburn, 1994).

(1) At the *founding stage*, a group of individuals (within an organization) gets together and agrees on the goal of developing a certain capability. To do so, "some leadership and mechanisms to govern the team" are required (Helfat & Peteraf, 2003, p. 1001).

(2) The *development stage* describes the incremental process of creating the capability by reviewing potential alternatives (based on the resources and capabilities already available) and pursuing the preferred capability development approach. Typically, the team either decides to build the capability from scratch or to imitate another organization's capability. Either way, "capability development entails improvement over time" through learning effects (Helfat & Peteraf, 2003, p. 1002).

(3) At the *maturity stage*, the capability is in place and needs to be maintained. "This involves exercising the capability, which refreshes the organizational memory" (Helfat & Peteraf, 2003, p. 1003).

Considering the dynamic nature of capabilities, their lifecycle is not linear. According to Helfat and Peteraf (2003), organizational capabilities are affected at the development and/or maturity stage by events taking place both inside and outside of the organization. These events may lead to a transformation of the organizational capability in six possible ways ("the six Rs of capability transformation", p. 1005), which may occur in parallel also: (1) retirement, (2) retrenchment, (3) replication, (4) renewal, (5) redeployment, or (6) recombination.

2.3.3 Data Management as a Dynamic Capability

Combining the understanding of data as a corporate resource with the RBV and the concepts of organizational capabilities, the author of this dissertation defines data management as a dynamic capability allowing enterprises to manage and provide the data resource with the aim of utilize it and generating business value (cf. Otto, 2012a, p. 14). As a dynamic capability, data management develops, combines, or reconfigures (data) resources as well as *core* and *ordinary data management capabilities*. In line with the concept of the capability lifecycle, data management is not a one-off effort, but an ongoing activity characterized by continuous improvement.

2.4 Reference Models and Maturity Models

Given the research objective of designing a capability reference model for strategic data management, which supports in structuring, designing, and assessing data management activities, this section clarifies the meaning of the terms *model*, *reference model*, and *maturity model*[18].

2.4.1 Models

A model is a simplified representation of a real-world object or situation, depicting the elementary aspects of it at a specific point in time (Becker, Delfmann, & Knackstedt, 2007; Schütte & Rotthowe, 1998). A model is the result of a design process and aims at depicting a complex matter in a way that it can be understood by its users (Frank, 2007). Examples of models in the IS domain are data models (representing an enterprise's core business objects as well as their characteristics and relations), business process models (depicting an enterprise's major activities and their sequence), or application architecture models (representing an enterprise's core applications as well as their functionalities and interfaces).

2.4.2 Reference Models

A reference model is a generic solution to a class of problems, which can be adjusted to a specific situation (Thomas, 2006; Vom Brocke, 2007). As a configurable "blueprint", it provides a reference point for designing a company-specific solution and increases the efficiency and effectiveness of the design process (Becker, Delfmann, Dreiling, Knackstedt, & Kuropka, 2004; Fettke & Loos, 2003a; Scheer & Nüttgens, 2000). A reference model results from a design process, in which a modeler specifies the critical elements of a system at a certain point in time (Schütte & Rotthowe, 1998). It is both descriptive and prescriptive as it allows for describing a certain domain, while at the same time offering guidance for designing this domain (Frank, 2007). According to Fettke and Loos (2003a), a reference model has three core characteristics: (1) it provides best practices, (2) it is universally applicable, and not designed to solve a single, isolated case, and (3) it is reusable, as it describes solution patterns which can be applied multiple times. Consequently, to apply a reference model in a real-world scenario, an enterprise

[18] Even though the research objective of this dissertation is the development of a reference model, many competing artifacts do not only provide guidance for designing data management (in the sense of a reference model) but also allow for analyzing the current status of data management (in the sense of a maturity model). Consequently, the term "maturity model" is introduced as well.

needs to adapt this model in a way that it suits its context and meets company-specific requirements (Fettke & Loos, 2003b).

As it is difficult to foresee all potential requirements and application scenarios when designing a configurable reference model, mechanisms for adaptation of the model (as general rules for reusing it in a specific scenario) are important to facilitate broader usage (Becker, Delfmann, & Knackstedt, 2004). To address the two major requirements of universal applicability and reusability, Vom Brocke (2007) introduces five adaptation mechanisms describing how elements of a reference model can be rearranged to develop a situation-specific model: (1) *analogy*, by creatively taking a reference model as an orientation for constructing a similar model, (2) *specialization*, by revising a reference model and extending its elements, (3) *aggregation*, by combining different elements of different reference models, (4) *instantiation*, by specifying generic elements of a reference model, and (5) *configuration*, by selecting elements from a variety of alternative choices. Configurable reference models include mechanisms for being adjusted based on context-specific (e.g. company size or industry) or application-specific configuration parameters (e.g. user group or use purpose) (Becker, Delfmann, Dreiling et al., 2004; Rosemann & Schütte, 1999).

Examples of reference models from the IS domain are the Architecture of Integrated Information Systems (ARIS) (Scheer & Nüttgens, 2000) or the Framework for Corporate Data Quality Management[19] (CDQM) (Otto, 2011b; Otto & Österle, 2015).

2.4.3 Maturity Models

A maturity model is a model specifically designed for organizational improvement, benchmarking, or self-assessment (Becker, Knackstedt, & Pöppelbuß, 2009; Mettler, Rohner, & Winter, 2010). It comprises a set of elements to describe the maturity of a design domain (Fraser, Moultrie, & Gregory, 2002). In general, *maturity* is defined as "the state of being complete, perfect or ready" (Oxford University Press, 2004). In the context of the IS domain, a maturity model is understood as the "description of the stages through which […] organizations evolve as they define, implement, measure, control, and improve their […] processes" (Paulk, Weber, Garcia, Chrissis, & Bush, 1993, p. A-

[19] The Framework for CDQM is explained in detail in *Subsection 3.2.1*.

4). The majority of maturity models adopts this definition and takes a process-oriented perspective (Ofner, Otto, & Österle, 2013)[20].

A maturity model consists of two sub-models: a domain model and an assessment model. "The domain model comprises criteria by which the design domain can be partitioned into discrete units to be assessed. The assessment model provides one or multiple assessment dimensions, each of which defining an assessment scale. What is basically assessed is to what extent certain criteria comply with the scale for each assessment dimension" (Ofner, Otto et al., 2013, p. 6).

Maturity models can be found in two forms:

(1) The *staged maturity model* assumes a certain sequence of stages along a desired development path in a design domain. Using such a model, assessments can be made as to what stage of development the design domain under analysis has reached (Becker et al., 2009; Paulk et al., 1993).

(2) The *continuous maturity model* assumes a dynamic development path (i.e. the development path is not predefined by the model). This type of maturity model is typically applied for reviewing a design domain on a regular basis by determining the maturity level of the different domain elements and identifying improvement potentials (EFQM, 2009).

One of the most prominent examples of a maturity model in the IS domain is the Capability Maturity Model (CMM) by the Carnegie Mellon Software Engineering Institute (SEI) (Paulk, 1997) (Mettler et al., 2010).

[20] Ofner et al. (2013) make the remark that the sole focus on the process is controversially discussed within the scientific community, and that a broader, more integrated view – covering also technological and organizational aspects – is demanded by many experts.

3. State of the Art: Reference Models and Maturity Models for Data Management

This chapter provides an overview of reference models and maturity models[21] used for data management, originating both from academia and practice. Its constituting sections outline the criteria underlying the author's decision to include or exclude certain models into/from his analysis (*Section 3.1),* describe and assess each model deemed relevant for this dissertation briefly in a separate subsection (*Section 3.2*), and present a comparative overview of the models selected (resulting from each model's evaluation in terms of meeting predefined design requirements), followed by implications derived from this evaluation to guide the research conducted for this dissertation (*Section 3.3*).

3.1 Criteria for Model Selection

Given the importance of data management in practice, and the growing interest in the topic in academia, a number of reference models (or "frameworks", as they are often called synonymously) and maturity models have been developed and proposed so far[22]. Domains of origin of these models are manifold, ranging from research, industry consortia, standardization bodies, and market analysts to consulting firms and software vendors (Pentek, Legner, & Otto, 2017). The following presentation and analysis focus on models of which the design is characterized either by scientific rigor or an effort of collaboration involving multiple subject matter experts from different organizational and thematic contexts, or both. This means that models developed and disseminated by consulting firms or software vendors are omitted from the analysis, as the author of the dissertation views these models to be single-expert and/or single-case induced, and set up mainly for marketing purposes[23]. In addition, the review considered only reference models and maturity models for data management that are state of the art and whose structure and components are publicly available and sufficiently detailed. *Table 3-1* gives an overview of the models presented and analyzed in the next section, including information such as model type, domain of origin, and year of publication.

[21] Although the research objective of this dissertation is the development of a reference model, the analysis of competing artifacts also considered maturity models as their underlying domain models can be regarded as reference models.

[22] Labadie and Legner (2017) provide a comprehensive overview of reference models for data management. However, they do not consider maturity models for data management.

[23] In his literature review, the author has reviewed also reference models and maturity models by consulting firms and software vendors. These are presented in *Appendix B*.

Table 3-1: Overview of data management models analyzed

No.	Model	Model type	Author(s) / organization of origin	Domain of origin	Year of publication
1	Framework for Corporate Data Quality Management	Reference model	CC CDQ; Schemm, J.; Otto, B.; Wende, K.; Schmidt, A.; Osl, P.; Kokemüller, J.; Weisbecker, A.; Gizanis, D.	Research, industry consortium	2007, 2011
2	Maturity Model for Enterprise Data Quality Management	Maturity model	CC CDQ; Ofner, M.; Hüner, K.; Otto, B.; Österle, H.	Research, industry consortium	2013, 2016
3	Capability Reference Model for Information Service Mgmt.	Reference model	CC CDQ; Bärenfänger, R.	Research, industry consortium	2017
4	Big Data Analytics Capability Model	Reference model	Akter, S.; Wamba, S.F.; Gunasekaran, A.; Dubey, R.; Childe, S.J.	Research	2016
5	Big Data Resources Framework	Reference model	Gupta, M.; George, J.F.	Research	2016
6	Master Data Management Maturity Model	Maturity model	Spruit, M.; Pietzka, K.	Research	2015
7	DAMA-DMBOK Framework	Reference model	DAMA	Industry consortium	2006, 2007, 2008, 2017
8	Data Quality Maturity Model	Maturity model	Performance Improvement Council (PIC)	Industry consortium	2016
9	Data Capability Assessment Model	Maturity model	EDM Council	Industry consortium	2018
10	IBM Data Governance Council Maturity Model	Maturity model	IBM Data Governance Council	Industry consortium, software vendor	2007
11	Data Quality Management System	Reference model	GS1	Standardization body	2010
12	Master Data Quality Management Framework	Reference model	ISO	Standardization body	2011
13	Data Management Capability Model	Reference model	Forrester Research, Inc.	Market analyst	2018
14	Enterprise Information Management Maturity Model	Maturity model	Gartner, Inc.	Market analyst	2014

3.2 Description and Assessment of Selected Models

3.2.1 Framework for Corporate Data Quality Management (CDQM)

The Framework for CDQM is a reference model for enterprise-wide management of master data (Otto & Österle, 2015). The reference model, which has its focus on data quality, was developed by researchers and practitioners of the CC CDQ at the University of St. Gallen based on Schemm's (2008) reference architecture for data synchronization. A first draft version of the model was presented by Otto et al. to the scientific community in 2007. The current version was published in 2011 (Otto, Kokemüller, Weisbecker, & Gizanis).

The basic structure of the Framework for CDQM follows the three-layer logic of the Business Engineering approach[24] (i.e. strategy, organization, and systems) (Österle & Winter, 2003). It defines six design areas for CDQM: (1) Strategy, (2) Controlling, (3) Organization, (4) Processes and Methods, (5) Architecture, and (6) Application Systems (see *Figure 3-1*). For specification of each design area, the CC CDQ has developed several methods, models, result types, and result documents (Otto & Österle, 2015).

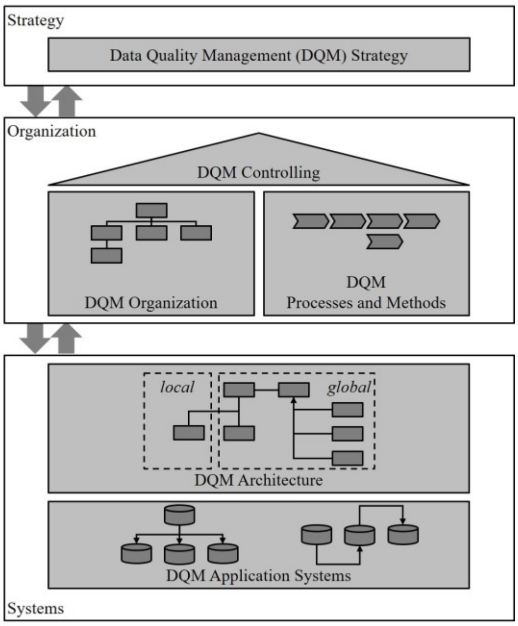


Figure 3-1: Framework for Corporate Data Quality Management (Otto & Österle, 2015, p. 24)

[24] Business Engineering is a systematic, step-wise approach for the digital transformation of enterprises with the help of specific methods and models (Österle & Winter, 2003; Österle, 2007).

Relevance of the model for this dissertation: The Framework for CDQM provided the basic structure for the research on quality-oriented data management in the CC CDQ for more than a decade. While it does not address the strategic importance of data, the six design areas it specifies are still relevant for data management in enterprises. Consequently, these design areas provided a baseline for the author of this dissertation for designing a capability reference model for strategic data management.

Table 3-2: Classification of the Framework for Corporate Data Quality Management

Model name	Framework for Corporate Data Quality Management				
Model type	*Reference model*		*Maturity model*		
Author(s) / Organization of origin	Schemm, J.; Otto, B.; Wende, K.; Schmidt, A.; Osl, P.; Kokemüller, J.; Weisbecker, A.; Gizanis, D. CC CDQ				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	2007 (first draft version) 2011 (current version)				

3.2.2 Maturity Model for Enterprise Data Quality Management (EDQM)

Based on the above described Framework for CDQM, Ofner, Otto et al. (2013) developed the Maturity Model for Enterprise Data Quality Management (EDQM). A first draft version of the maturity model was presented to the scientific community in 2009 (Hüner, Ofner, & Otto), while the first practice-oriented version was published in 2011 (EFQM, 2011). In 2013, the final version of the model was presented to the scientific community (Ofner, Otto et al.). The practitioners' version was updated in 2016 (EFQM, 2016).

The Maturity Model for EDQM is a continuous maturity model for DQM, which adopts the Enablers-Results logic of the EFQM Excellence Model[25]. On a micro-level, it

[25] The EFQM Excellence Model is a non-prescriptive framework for developing a culture of excellence within enterprises. The model comprises nine criteria, which are subdivided into "enablers" and "results". "Enablers refer to what an organization does, whereas results refer to what an organization achieves" (EFQM (2016, p. 15). Enablers include five criteria, i.e. (1) leadership, (2) people, (3) strategy, (4) partnerships & resources, and (5) processes, products & services, while results include four criteria, i.e. (6) people, (7) customer, (8) society, and (9) business.

specifies 30 practices and 56 measures that can be used for concrete maturity assessments. On a macro-level, the model adopts the six design areas of the Framework for CDQM as Enablers and puts them in relation with the four Result criteria of the EFQM Excellence Model. The Maturity Model for EDQM has been applied in more than 70 projects so far (Pentek, 2017).

Relevance of the model for this dissertation: The Maturity Model for EDQM is based on the Framework for CDQM and has been applied in multiple enterprises. As the author of this dissertation aimed at developing a reference model of high utility, the design approach as well as the structure, practices, and measures of this maturity model served as an important basis for the author to design the capability reference model and specify its success criteria and recommended practices.

Table 3-3: Classification of the Maturity Model for Enterprise Data Quality Management

Model name	Maturity Model for Enterprise Data Quality Management				
Model type	*Reference model*			*Maturity model*	
Author(s) / Organization of origin	Ofner, M.; Hüner, K.; Otto, B.; Österle, H. CC CDQ, EFQM				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	2009: First draft version for scientific community (Hüner et al.) 2011: First practice-oriented version (EFQM) 2013: Full version for scientific community (Ofner, Otto et al.) 2016: Updated practitioners' version (EFQM)				

3.2.3 Capability Reference Model for Information Service Management

Bärenfänger's Capability Reference Model for Information Service Management is another outcome of the research conducted within the CC CDQ. Published in 2017, it provides a reference for designing and implementing information service management in enterprises (Bärenfänger, 2017). With its focus on information services, which are defined as content-providing services being "the primary value-creating outputs of information management" (Bärenfänger, 2017, p. 32), the reference model has a broader scope than data management. It defines three main capability groups: (1) information service planning and management capabilities, (2) technical information service (IS value chain) capabilities, and (3) information service prerequisite capabilities (see *Table 3-4*). Furthermore, the model distinguishes seven capabilities, 26 sub-capabilities, and 84 elementary capabilities (Bärenfänger, 2017).

Table 3-4: Capability Reference Model for Information Service Management (Bärenfänger, 2017, p. 156)

Main capability group	Capability	Sub-capability
Information service (IS) planning and management capabilities	Business relationship management	– Strategic alignment – Governance – Business capability understanding
	Information service management	– IS portfolio management – IS portfolio integration – IS controlling
	IS lifecycle management	– IS specification – IS design and development – IS change and maintenance
Technical information service (IS) (IS value chain) capabilities	Data processing / IS value chain	– Data acquisition – Data preprocessing – Data transformation – IS delivery – Data management – Reporting – Analytics
	IS architecture and operation	– Cross-service interoperability – Quality-of-service assured storage and processing – IS administration
Information service (IS) prerequisite capabilities	Core BO-specific information management	– Product information management – Customer information management – Ecosystem information management – Performance information management

Main capability group	Capability	Sub-capability
	Data governance and data quality management (DQM)	– Data architecture and DQM – Organization and process management – Application management

Relevance of the model for this dissertation: The Capability Reference Model for Information Service Management applies the RBV to information/data management. Furthermore, it provides a comprehensive list of capabilities which are (in part) relevant for data management. Moreover, it explicitly emphasizes the strategic importance of data management, as well as the necessity to incorporate the business view and consider the impact on business when planning and managing data-related capabilities. Especially the linkage between data management and business (i.e. the business relationship management capability), the portfolio view (i.e. the information service management capability) as well as the model's orientation towards capabilities had an influence on the design activities conducted for this dissertation.

Table 3-5: Classification of the Capability Reference Model for Information Service Management

Model name	Capability Reference Model for Information Service Management				
Model type	*Reference model*		*Maturity model*		
Author(s) / Organization of origin	Bärenfänger, R. CC CDQ				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	2017				

3.2.4 Big Data Analytics Capability (BDAC) Model

The Big Data Analytics Capability (BDAC) Model by Akter, Wamba, Gunasekaran, Dubey, and Childe unites relevant capabilities for managing big data and conducting big data analytics based on an extensive literature review (Akter et al., 2016). It presents these capabilities in a hierarchical reference model consisting of three primary capability dimensions (i.e. management, technology, and talent) and eleven sub-dimensions (see *Table 3-6*).

Table 3-6: Big Data Analytics Capability Model (Akter et al., 2016, p. 121)

Primary dimensions	Sub-dimensions
BDA management capability	– BDA planning – BDA investment – BDA coordination – BDA control
BDA technology capability	– BDA connectivity – BDA compatibility – BDA modularity
BDA talent capability	– BDA technology management knowledge – BDA technical knowledge – BDA business knowledge – BDA relational knowledge

Relevance of the model for this dissertation: Like Bärenfänger's Capability Reference Model for Information Service Management, the BDAC Model provides a structured list of relevant capabilities for managing big data and conducting big data analytics. The reference model had an influence on the design activities conducted for this dissertation, as it underlines the need for strategic data management in the light of analytics activities (i.e. the BDA management capability) and highlights the importance of well-trained and well-informed employees (i.e. the BDA talent capability).

Table 3-7: Classification of the Big Data Analytics Capability Model

<table>
<tr><td>Model name</td><td colspan="5">Big Data Analytics Capability (BDAC) Model</td></tr>
<tr><td>Model type</td><td colspan="3">Reference model</td><td colspan="2">Maturity model</td></tr>
<tr><td>Author(s) / Organization of origin</td><td colspan="5">Akter, S.; Wamba, S.F.; Gunasekaran, A.; Dubey, R.; Childe, S.J.</td></tr>
<tr><td>Domain of origin</td><td>Research</td><td>Industry consortium</td><td>Standardization body</td><td>Market analyst</td><td>Consulting firm / Software vendor</td></tr>
<tr><td>Year of publication</td><td colspan="5">2016</td></tr>
</table>

3.2.5 Big Data Resources Framework

In their Big Data Resources Framework, Gupta and George (2016) define seven resources, which, combined with each other, allow building a big data analytics capability. The authors assign each of the seven resources to one of three classes (i.e. tangible, human, and intangible) (see *Table 3-8*). For each resource, they introduce between two and six items (32 in total), which specify the requirements to be met by the respective resource from a big data analytics point of view (Gupta & George, 2016).

Table 3-8: Big Data Resources Framework (Gupta & George, 2016, p. 1051)

Class	Resource
Tangible big data resources	– Data (internal, external, merging internal and external) – Technology (e.g. Hadoop, NoSQL) – Basic resources (i.e. time, investment)
Human big data resources	– Managerial skills (i.e. analytics acumen) – Technical skills (i.e. educational and trainings pertaining to big data-specific skills)
Intangible big data resources	– Data-driven culture (i.e. preference for decisions based on data rather than on intuition) – Intensity of organizational learning (i.e. ability to explore, store, share, and apply knowledge)

Relevance of the model for this dissertation: The Big Data Resources Framework adds a classification of resources to the capability view of strategic data management and emphasizes the importance of people- and culture-oriented aspects. Even though considering data as a “tangible” resource is questionable, the model’s focus on the availability of required resources (i.e. time and investment), on education and training, and on cultural aspects of data management had an influence on the artifact designed for this dissertation.

Table 3-9: Classification of the Big Data Resources Framework

Model name	Big Data Resources Framework				
Model type	*Reference model*			*Maturity model*	
Author(s) / Organization of origin	Gupta, M.; George, J.F.				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	2016				

3.2.6 Master Data Management Maturity Model (MD3M)

The Master Data Management Maturity Model (MD3M) by Spruit and Pietzka is a staged maturity model focusing on master data. Its domain model covers five key topics: (1) data model, (2) data quality, (3) usage and ownership, (4) data protection, and (5) maintenance (see *Table 3-10*). Each key topic comprises a number of focus areas (13 in total), which can be described on five maturity levels, resulting in a list of 65 capability stages (Spruit & Pietzka, 2015).

Table 3-10: Master Data Management Maturity Model (Spruit & Pietzka, 2015, p. 1072)

Key topics	Focus areas
(1) Data model	– Definition of master data – Master data model – Data landscape
(2) Data quality	– Assessment of data quality – Impact on business – Awareness of quality gaps – Improvement
(3) Usage and ownership	– Data usage – Data ownership – Data access
(4) Data protection	– Data protection

Key topics	Focus areas
(5) Maintenance	– Storage – Data lifecycle

Relevance of the model for this dissertation: Despite the model's sole focus on master data, the author of this dissertation found the detailed descriptions of the 65 capability stages (particularly on data access and data protection) useful. Actually, all 13 focus areas of the MD3M are integrated (as success criteria or recommended practices) in the resulting artifact of this dissertation.

Table 3-11: Classification of the Master Data Management Maturity Model

Model name	Master Data Management Maturity Model (MD3M)				
Model type	*Reference model*			*Maturity model*	
Author(s) / Organization of origin	Spruit, M.; Pietzka, K.				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	2015				

3.2.7 DAMA-DMBOK Framework

The Data Management Association (DAMA) is the most notable, global association of data management professionals (Otto, 2012a). It has issued the "Data Management Body of Knowledge (DMBOK)" as an industry reference for data management practices (DAMA, 2009), which was recently published in its second edition as "DMBOK2" (DAMA, 2017), and the accompanying "Dictionary of Data Management" (DAMA, 2008b, DAMA, 2011). At the heart of these publications is the DAMA-DMBOK Framework, which has undergone an evolution comprising several versions since its first publication in 2006 (DAMA, 2008a, DAMA, 2009, DAMA, 2017). The current version – called "DAMA-DMBOK2 Data Management Framework (The DAMA Wheel)" – comprises eleven knowledge areas (see *Figure 3-2*). "It places data governance at the center of data management activities, since governance is required for consistency within and balance between the functions. The other Knowledge Areas are balanced around the wheel. They are all necessary parts of a mature data management function, but they may be implemented at different times, depending on the requirements of the organization" (DAMA, 2017, p. 35).

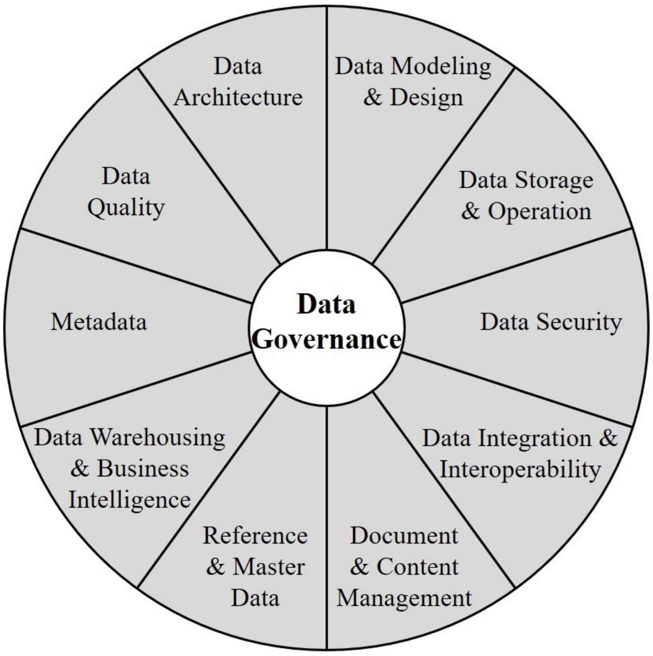


Figure 3-2: DAMA-DMBOK2 Data Management Framework (DAMA, 2017, p. 36)

Each knowledge area comes with a list of activities (summing up to 102 activities in total), and each activity is to be executed along four phases: (1) Plan, (2) Develop, (3) Control, and (4) Operate. Following this procedure ensures structured, measurable, and result-oriented implementation and improvement of each knowledge area (DAMA, 2017).

In addition to the DAMA-DMBOK Framework, DAMA has introduced the "Environmental Factors Hexagon", indicating the close link between people, processes, and technology in data management (see *Figure 3-3*). Its components are to be reflected within each knowledge area of the DAMA-DMBOK Framework. "It puts goals and principles at the center, since these provide guidance for how people should execute activities and effectively use the tools required for successful data management" (DAMA, 2017, p. 35).

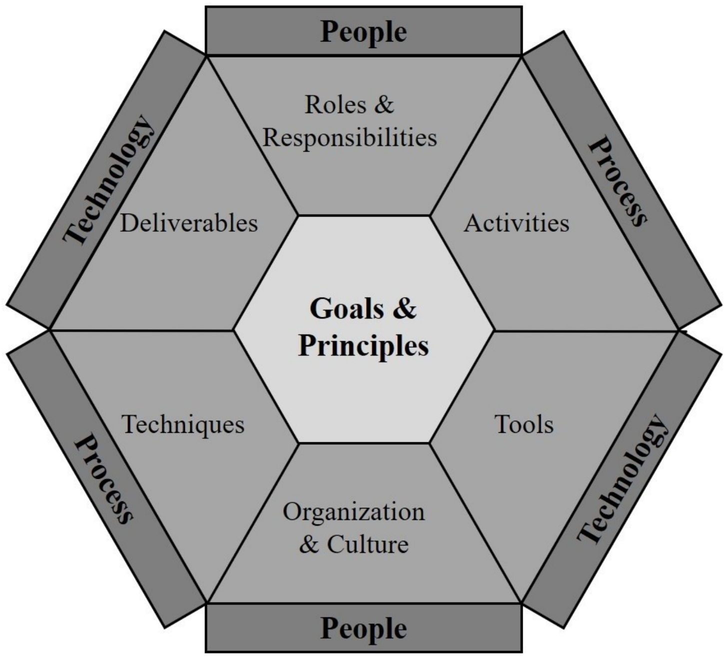


Figure 3-3: DAMA Environmental Factors Hexagon (DAMA, 2017, p. 36)

Relevance of the model for this dissertation: Given the reputation of DAMA as the most relevant, global practitioners' association in the data management domain, the widespread adoption of its recommendations by practitioners, and the frequent updates of DAMA's publications and models, the DAMA-DMBOK Framework can be regarded as the state-of-the-art reference model for data management in the practitioners' community. In addition to its practical relevance, DAMA's reference model addresses most of the requirements on strategic data management (as introduced in *Subsection 2.2.5*).

The DMBOK2 (2017) includes a chapter entitled "Big Data and Data Science" (pp. 497-528). However, DAMA does not consider big data analytics as an element of its reference model, leading to a disconnect between data analytics and the eleven knowledge areas of the DAMA-DMBOK Framework. Furthermore, the DMBOK2 (2017) contains a chapter named "Data Management Maturity Assessment" (pp. 531-549), which does not introduce a maturity assessment based on the reference model, but provides an overview of maturity models from other sources (which, however, do not fully reflect the previously introduced knowledge areas). Consequently, the DAMA-DMBOK Framework's content, but not its structure, had an influence on the design process conducted for this dissertation.

Table 3-12: Classification of the DAMA-DMBOK Framework

<table>
<tr><td>Model name</td><td colspan="10">DAMA-DMBOK Framework</td></tr>
<tr><td>Model type</td><td colspan="5">Reference model</td><td colspan="5">Maturity model</td></tr>
<tr><td>Author(s) / Organization of origin</td><td colspan="10">Data Management Association International (DAMA)</td></tr>
<tr><td>Domain of origin</td><td colspan="2">Research</td><td colspan="2">Industry consortium</td><td colspan="2">Standardization body</td><td colspan="2">Market analyst</td><td colspan="2">Consulting firm / Software vendor</td></tr>
<tr><td>Year of publication</td><td colspan="10">2006: Version 1.0 of the DAMA-DMBOK Functional Framework
2007: Version 2.0 of the DAMA-DMBOK Functional Framework
2008: First edition of the DAMA Dictionary of Data Management
2008: Version 3.0 of the DAMA-DMBOK Functional Framework
2009: First edition of DAMA-DMBOK
2011: Second edition of the DAMA Dictionary of Data Management
2017: Second edition of DAMA-DMBOK
2017: DAMA-DMBOK2 Data Management Framework</td></tr>
</table>

3.2.8 Data Quality Maturity Model

The Performance Improvement Council (PIC) is a US-based consortium of governmental agencies pursuing the objective of exchanging best practices. Its Data Quality Working Group developed the Data Quality Maturity Model, which is a staged model considering four elements: (1) policies and procedures, (2) quality control and assurance practices, (3) governance and leadership (including culture), and (4) human capital (PIC, 2016). For each element, the model roughly specifies four maturity stages and provides short recommendations for proceeding to the next stage of maturity.

Relevance of the model for this dissertation: The model emphasizes the importance of data quality for all kinds of organizations (also governmental agencies). While the model and its four elements provide an incomplete view on data management, the descriptions of the highest maturity level of each element provide valuable input with regard to policies, data quality applications, roles, leadership, training, and culture.

Table 3-13: Classification of the Data Quality Maturity Model

Model name	Data Quality Maturity Model				
Model type	*Reference model*		*Maturity model*		
Author(s) / Organization of origin	Performance Improvement Council (PIC)				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	2016				

3.2.9 Data Capability Assessment Model (DCAM)

The Enterprise Data Management (EDM) Council is an industry association for financial services. The EDM Council's Data Capability Assessment Model (DCAM) is a staged maturity model, the domain model of which comprises eight components (see *Table 3-14*).

Table 3-14: Components of the Data Capability Assessment Model (EDM Council, 2018, p. 5)

Component	Component description
(1) Data management strategy	– long-term goal of the data management program – blueprint to gain internal alignment among stakeholders and define how the organization will approach the management of data content
(2) Data management business case	– justification of the data management program – mechanism for ensuring sufficient and sustainable funding – approach for measuring the costs and benefits of EDM
(3) Data management program	– mechanism for EDM implementation – stakeholder engagement – communications program and education on the concepts of data content management – engagement model and operational routines
(4) Data governance	– rules of engagement for implementation of the data management program – focus is on the implementation of policies, standards, and operational procedures necessary to ensure that stakeholders "behave"

Component	Component description
(5) Data architecture	– "design of information content" including: – identification of data domains, – establishment of taxonomies, – alignment with contractual obligations, – documentation of metadata, and – designation of CDEs
(6) Technology architecture	– "design of physical architecture" including the platforms and tools in support of data management implementation – definition of how data is acquired, stored, integrated and distributed
(7) Data quality	– provide data to business users that is fit-for-purpose – provide data that users trust and have confidence in (in terms of being exactly what they expected, without the need for reconciliation and data transformation)
(8) Data control environment	– integrate components into a cohesive operational model – ensure that control mechanisms are in place to achieve consistency across the entire data lifecycle – align with organizational privacy and security policies

These components are further subdivided into 36 capabilities, 112 sub-capabilities, and 306 objectives (EDM Council, 2018). Each sub-capability is documented by a one-sentence statement, a short description, a specification of its objectives, an advice from an audit perspective, guiding questions, a list of artifacts (giving proof of the existence of the sub-capability), and a scoring guide (describing the six maturity stages of the sub-capability).

Relevance of the model for this dissertation: The author of the dissertation has regarded the balanced, hierarchical structure of DCAM, the comprehensive list of capabilities and sub-capabilities, and the detailed documentation of its elements as highly relevant for the research he conducted for this dissertation. Hence DCAM documentation served as an important reference in the model design process. From a content perspective, DCAM provides a comprehensive list of relevant data management aspects – including a business case for data management.

Table 3-15: Classification of the Data Management Capability Assessment Model

Model name	Data Management Capability Assessment Model				
Model type	*Reference model*		*Maturity model*		
Author(s) / Organization of origin	EDM Council				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	2018				

3.2.10 IBM Data Governance Council Maturity Model

The IBM Data Governance Council is an association of 55 organizations (mainly from the financial industry but also including three universities). Between 2004 and 2007, this association developed the IBM Data Governance Council Maturity Model. The staged maturity model, which is based on the SEI's CMM, is described as "a tool to assess your organization's current state of Data Governance awareness and effectiveness. Through this concrete and objective set of benchmarks, organizations can evaluate current gaps in their data governance practices and define new opportunities quickly for improving the governance based upon the observable behaviors and insights of the Data Governance Council" (IBM Data Governance Council, 2007, p. 14). The maturity model's domain model defines four groups (Enablers, Outcomes, Core Disciplines, and Supporting Disciplines), each of which comprises a number of data governance domains (eleven in total) (see *Table 3-16*).

Table 3-16: IBM Data Governance Council Maturity Model (IBM Data Governance Council, 2007, 8-10)

Group	Domain	Domain description
Enablers	Organizational Structures & Awareness	Level of mutual responsibility between business and IT, and recognition of the fiduciary responsibility to govern data at different levels of management
	Stewardship	Quality control function designed to ensure custodial care of data for asset enhancement, risk mitigation, and organizational control
	Policy	Written articulation of desired organizational behavior

Group	Domain	Domain description
Out-comes	Value Creation	Process by which data assets are qualified and quantified to enable business to maximize the value created by data assets
	Data Risk Management & Compliance	Methodology by which risks are identified, qualified, quantified, avoided, accepted, or mitigated
Core Disciplines	Data Quality Management	Methods to measure, improve and certify the quality and integrity of production data, test data, and archived data
	Information Security & Privacy	Policies, practices, and control mechanisms used by an organization to mitigate risk and protect data assets
	Information Lifecycle Management	Systematic, policy-based approach for information collection, use, retention, and deletion.
Supporting Disciplines	Data Architecture	Architectural design of structured and unstructured data systems and applications making data available to users
	Classification & Metadata	Methods and tools to create common semantic definitions of business and IT terms, data models, data types, and data repositories; metadata to bridge the gap between human and computer understanding
	Audit Information, Logging & Reporting	Organizational processes for monitoring and measuring the data value, risks, and the efficacy of governance

Relevance of the model for this dissertation: With only eleven elements on its micro-level, the model does not provide a sufficiently detailed perspective on data management. However, by adopting the enablers-outcomes view, the maturity model promotes a business-oriented results perspective on data management, which had an influence on the structure and logic of the artifact developed by the author of this dissertation. Furthermore, the model addresses additional data-related concerns, such as data security and data privacy, which the author also reflected in his design activities.

Table 3-17: Classification of the IBM Data Governance Council Maturity Model

Model name	IBM Data Governance Council Maturity Model				
Model type	*Reference model*		*Maturity model*		
Author(s) / Organization of origin	IBM Data Governance Council				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	2004 (first version) 2007 (second and latest version)				

3.2.11 Data Quality Management System

In 2010, GS1, a not-for-profit organization that develops and maintains global standards for business communication, issued the third version of its Data Quality Framework. It consists of the Data Quality Management System (DQMS), a self-assessment tool based on the DQMS, a product inspection procedure, and a reference documentation. The DQMS "provides guidance for organisations [sic] to establish, implement, maintain and improve a series of processes and activities related to the management of information and data quality of their master data output" (GS1, 2010, p. 7). The DQMS is structured as a matrix with four activity types in the vertical direction, which are called (1) Plan, (2) Document, (3) Execute, and (4) Monitor, and four aspects in the horizontal direction, namely (1) Organizational Capabilities, (2) Policies & Standards, (3) Business Processes, and (4) Systems Capabilities (see *Table 3-18*). For each field of the matrix, several capabilities are defined (73 in total). Each capability is specified by an explanation, providing a short description, the capability's rationale, its target group, an implementation example, and relevant questions to ask when assessing the capability.

Table 3-18: Data Quality Management System (GS1, 2010, p. 11)

	Organizational Capabilities	**Policies & Standards**	**Business Processes**	**System Capabilities**
Plan	– Executive sponsorship (mission & vision) – Accountable leadership – Staff roles & skill set – Data owners & stakeholders – Data governance office	– Mission & vision – Goals & objectives – Guiding principles – Success measures – Action plans – Policy & standards management	– Initial data entry & setup – Ongoing data maintenance – Processes involved in the information's lifecycle	– Unified data repository – Design & architecture – Workflow & user interface – Data validation – Security & access control – Revision & change history – External publication – Internal publication
Document	– Governance organizational structure – Roles & responsibilities – Personal objectives – Reporting alignment	– Mission, goals, principles and success measures – Governance model & decision process – Data definitions & standards – Security & use policy – Audit procedures – Documentation standards – Risk management – Customer feedback policy	– Operating procedures – Process flow diagrams – Job aids & work instructions – Performance metrics	– System requirements – Operating procedure – Performance metrics
Execute	– Education & awareness – Internal communication – Training	– Education & awareness – Documentation management – Policies & standards management – Data issue management – Training – Customer feedback resolution	– Education & awareness – Performance management – Process issue management – Change management	*None*

	Organizational Capabilities	Policies & Standards	Business Processes	System Capabilities
Monitor	– Organizational capability review – Review of personal objectives	– Policy & standards review	– Workflow controls – System validation – Performance reporting on service levels – Performance reporting on data quality – External & internal feedback – Process compliance audits – Product measurements – Review & reporting of audit results – Monitor impact of erroneous data	– Performance reporting on service levels

Relevance of the model for this dissertation: The GS1 model uses a matrix structure for putting management activities (i.e. plan, document, execute, monitor) and organizational, procedural, and technological capabilities in relation. Particularly the definition and use of the four management activities had an influence on the structure of the artifact to be developed for the dissertation. Furthermore, the detailed documentation of quality-oriented management capabilities provided the basis for developing and documenting the success criteria and recommended practices of the capability reference model for strategic data management.

Table 3-19: Classification of the Data Quality Management System

Model name	Data Quality Framework				
Model type	*Reference model*			*Maturity model*	
Author(s) / Organization of origin	GS1				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	*year of publication of the previous two version unknown* 2010 (third version)				

3.2.12 Master Data Quality Management Framework

In 2011, the standardization body ISO published ISO 8000 standard, which "specifies fundamental principles of master data quality management, and requirements for implementation, data exchange and provenance. This standard also contains an information framework that identifies processes for data quality management" (International Organization for Standardization [ISO], 2011, p. vi). The framework defines three top-level processes, each of which containing three lower-level processes (see *Table 3-20*). In addition, ISO 8000 defines three roles (i.e. data manager, data administrator, data technician). Each lower-level process is defined by a process description, a list and explanation of its constituting activities, a role responsible for it (including a specification of tasks assigned to the role), and a documentation of the relationship of the process to other processes in the framework.

Table 3-20: Master Data Quality Management Framework (ISO, 2011, pp. 7–8)

Top-level process	Lower-level process	Lower-level process description
Data operations	Data architecture management	Manages the organization-wide data architecture from an integrated perspective to use data consistently in distributed information systems and thereby ensure data quality
	Data design	Designs data schema and implements a database to enable data users to use data correctly and thereby ensure data quality
	Data processing	Creates, searches for, updates and deletes data in accordance with guidelines of data operations
Data quality monitoring	Data quality planning	Sets up data quality objectives in compliance with the strategy of the organization; identifies factors to be managed; performs actions in order to accomplish objectives (includes also data quality assurance and subsequent adjustment of data quality objectives)
	Data quality criteria setup	Sets up criteria with regard to characteristics of data and measuring methods
	Data quality measurement	Measures target data against the criteria set up in the previous process (in real-time or periodically)
Data quality improvement	Data stewardship / flow management	Analyzes data operations and data flows between organizations; identifies responsible parties and data operation systems including their influence on data quality; manages the stewardship of data operations
	Data error cause analysis	Analyzes root causes of data errors and prevents recurrence of the same errors
	Data error correction	Corrects defective data

Relevance of the model for this dissertation: With capabilities being operationalized in processes, ISO's process framework provides an overview of relevant data-quality oriented capabilities, which are specified on a reasonable level of detail. Furthermore, the framework introduces (few) relevant roles and their responsibilities. However, as the framework focuses on master data and data quality only, the author of the dissertation used it mainly as an orientation with regards to the documentation structure and level of detail.

Table 3-21: Classification of the Master Data Quality Management Framework

Model name	Master Data Quality Management Framework				
Model type	*Reference model*		*Maturity model*		
Author(s) / Organization of origin	International Organization for Standardization (ISO)				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	2011				

3.2.13 Data Management Capability Model

The Data Management Capability Model by market analyst Forrester Research considers data management as a set of capabilities covering aspects of people, processes, and technology (Hopkins, Leganza, Goetz, & Lee, 2018). The reference model defines three basic value streams, each of which comprises a number of capabilities (nine in total) (see *Table 3-22*).

Table 3-22: Data Management Capability Model (Hopkins et al., 2018)

Value stream	Capability
Data management planning and data architecture development	– Data architecture development – Data management technology research and planning – Data innovation R&D
Service delivery	– Business data services delivery

Value stream	Capability
	– Data management technology implementation and maintenance – Data management technology operation
Security and governance	– Data security – Data governance
Across all three value streams	– Stakeholder engagement

Relevance of the model for this dissertation: Although the model is rather generic and defines only a limited number of capabilities, some of the capabilities it contains (such as data innovation R&D or technology research and planning) are unique across the range of models reviewed. The author of the dissertation considered these capabilities in his research activities, while the model as a whole was less important for the research conducted.

Table 3-23: Classification of the Data Management Capability Model

Model name	Data Management Capability Model				
Model type	*Reference model*		*Maturity model*		
Author(s) / Organization of origin	Forrester Research, Inc.				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	2018				

3.2.14 Enterprise Information Management (EIM) Maturity Model

Market analyst Gartner published the initial version of its Enterprise Information Management (EIM) Maturity Model in 2008 (Newman & Logan, 2008). The latest version was released in 2014. It is a staged maturity model, whose domain model comprises seven building blocks (see *Table 3-24*). For each building block, the model defines certain stages of maturity and proposes action items to improve maturity (Gartner, 2014).

Table 3-24: Enterprise Information Management Maturity Model (Gartner, 2014)

Building block	Objective
Vision	EIM vision that enables the organization's business vision
Strategy	EIM strategy and roadmap, based on EIM, business needs, and the status quo
Metrics	EIM business case, based on improvements or attainment of business outcomes
Information governance	Information governance framework with clearly defined responsibilities and accountabilities
Organization and roles	Necessary structures and roles to support information strategy, governance, and stewardship
Information lifecycle	Documentation and understanding of the flow of information across the organization to optimize business processes, governance, and organization
Information infrastructure	Principles to rationalize and modernize information management-related tools and technology to maximize information asset utilization

Relevance of the model for this dissertation: As Gartner's maturity model is rather generic, lacks a sufficient level of detail, and defines only a limited number of building blocks, the author considered it to be less relevant for his overall research activities. However, the strategic view of data, which the model suggests, and its emphasis on the need for strategic planning and business-orientation, influenced the logic of the reference model developed by the author.

Table 3-25: Classification of the Enterprise Information Management Maturity Model

Model name	Gartner Enterprise Information Management (EIM) Maturity Model				
Model type	*Reference model*		*Maturity model*		
Author(s) / Organization of origin	Gartner, Inc.				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	2008 (first version) 2014 (second and last version)				

3.3 Comparative Overview and Implications for the Dissertation

As this dissertation focuses on the development of a capability reference model for strategic data management, the models presented in the previous subsection are evaluated against the distinctive features of strategic data management (resulting from the author's literature review and summarized in *Table 2-2* in *Subsection 2.2.5*). These distinctive features basically relate to (1) the role of data, (2) the management objective, (3) data-related concerns, and (4) data sources. Furthermore, the evaluation reviews whether the examined models include (5) constituents on strategic, organizational, and technological level:

(1) Role of data: Data, being a strategic resource, must enable (1a) operational excellence (i.e. efficient business processes and decision-making), (1b) new/enhanced business models (i.e. data-enriched products/services and/or data-driven services), and (1c) risk reduction and compliance.

(2) Data management objective: Strategic data management must generate business value.

(3) Data-related concerns: Strategic data management must target a number of concerns, such as (3a) data quality, (3b) compliance, (3c) data privacy, and (3d) data security.

(4) Data sources: Strategic data management must address (4a) internal data sources and (4b) external data sources.

(5) Constituents: Strategic data management, being a socio-technological design capability, must cover strategic (4a), organizational, and technological aspects (4c).

Table 3-26 presents a comparative overview of the models, resulting from each model's evaluation in terms of addressing each of the above defined requirements (i.e. "requirement fully addressed", "requirement partially addressed", and "requirement not addressed").

Table 3-26: Model evaluation and comparative overview

	(1a) Data as an enabler of operational excellence	(1b) Data as an enabler of new/enhanced business models	(1c) Risk reduction and compliance	(2) Business value orientation	(3a) Data quality	(3b) Compliance	(3c) Data privacy	(3d) Data security	(4a) Internal data sources	(4b) External data sources	(5a) Strategic constituents	(5b) Organizational constituents	(5c) Technologic constituents
Framework for Corporate Data Quality Management	X	-	-	(x)	X	-	-	-	X	-	X	X	X
Maturity Model for Enterprise Data Quality Management	X	-	-	(x)	X	-	-	-	X	-	X	X	X
Capability Reference Model for Information Service Mgmt.	X	X	(x)	(x)	X	(x)	X	X	X	X	X	(x)	X
Big Data Analytics Capability Model	X	X	-	X	(x)	-	-	-	X	X	X	X	X
Big Data Resources Framework	X	X	-	-	-	-	-	-	X	X	(x)	X	X
Master Data Management Maturity Model	X	-	-	X	X	-	(x)	X	X	-	-	X	X
DAMA-DMBOK Framework	X	(x)	X	(x)	X	X	X	X	X	(x)	X	X	X
Data Quality Maturity Model	X	-	X	(x)	X	X	-	-	X	-	-	X	(x)
Data Capability Assessment Model	X	-	X	X	X	X	X	X	X	-	X	X	X
IBM Data Governance Council Maturity Model	(x)	(x)	X	X	X	X	X	X	X	(x)	(x)	X	X
Data Quality Management System	X	-	(x)	(x)	X	(x)	(x)	X	X	X	X	X	X
Master Data Quality Management Framework	X	-	-	(x)	X	-	-	-	X	X	-	X	X
Data Management Capability Model	(x)	(x)	-	(x)	-	-	-	X	X	(x)	(x)	X	X
Enterprise Information Management Maturity Model	(x)	(x)	-	X	X	-	-	-	X	X	X	X	X
X: fully addressed; (x) partially addressed; -: not addressed													

From the analysis and assessment of the models investigated, both individually and in comparison with each other, the author of the dissertation derived the following implications for his research:

– All reference models and maturity models under analysis regard data (at least implicitly) as an enabler of operational excellence. However, only few models reflect the strategic role of data as an enabler of data-driven business models, products, and services. Even fewer propose a strategic approach for data management, while aspects of digital business transformation and the data-driven enterprise are addressed only rudimentarily. Consequently, the research undertaken for this dissertation aimed at emphasizing the enabling role of data (regarding multiple, business-critical purposes) and the strategic character of data management.

– While many publishers (i.e. authors and organizations the models originate from) point to the business relevance of data management, only few models address this idea explicitly by incorporating elements to create transparency regarding the output and business impact of data management. Therefore, the artifact developed for this dissertation is supposed to explicitly link data management and its results with business.

– In line with the general scientific view presented in *Section 2.2*, the majority of the models investigated consider high-quality data as the most important goal of data management. Only half of the models consider aspects like compliance, data privacy, or data security. Furthermore, most – and even the most recent – models limit their scope to (internal) master data, neglecting the relevance of further data types and external data sources. The research undertaken for this dissertation aimed at closing this gap by explicitly considering all data types – originating both from internal and external sources – as relevant for data management.

– While almost all models cover organizational and technological aspects of data management implementation, the strategic dimension is largely neglected. To close that gap, the research conducted for this dissertation aimed at establishing a comprehensive view of data management covering organizational, technological, *and* strategic aspects.

– Only Bärenfänger's Capability Reference Model for Information Service Management (2017) and the DAMA-DMBOK Framework (2017) address all requirements of strategic data management (fully or partially). However, with its focus on information services, Bärenfänger's model has a broader scope, elaborating data management on a less detailed level as deemed necessary by the author of this dissertation for his research goal. The DAMA model, on the other hand, does show a considerable level of detail, but has deficits due to its knowledge areas not being linked with each other, and the important topics of big data and data analytics not being integrated in the model.

– Generally, the number of related works presented (in the previous section and in the appendix) indicates the importance of data management for various groups – such as researchers, industry associations, standardization bodies, market analysts, consulting firms, and software vendors – and the practical relevance of structuring its constituents. Most of the models investigated originated from the practitioners' domain. These models typically lack the scientific rigor – following guidelines of orderly modeling, for example – in contrast to models developed by researchers. Some of the practitioners' models rather appear as a collection of experiences and best practices, with the structure and hierarchy of elements being somewhat imbalanced and the links between single model elements remaining unclear. Nevertheless, especially standardization bodies have come up with very comprehensive data management models often featuring a detailed documentation of model elements. To advance the state of the art, the author of the dissertation put great emphasis on providing a comprehensive, well-balanced, consistent, and detailed artifact resulting from a rigorous, scientifically accepted design process.

As the review of competing artifacts shows, none of the models investigated fully meets the requirements of strategic data management. This finding substantially motivated the research conducted for this dissertation.

4. Research Design

This chapter presents the research design underlying the construction of the capability reference model as the key artifact of this dissertation. *Section 4.1* outlines the overall research methodology and individual research methods applied. *Section 4.2 i*ntroduces the CC CDQ as the consortium research program which the research conducted for this dissertation was embedded in and which influenced the author's design decisions through the knowledge and experiences accumulated since its foundation in 2006. *Section 4.3* then details the concrete research activities conducted for this dissertation.

4.1 Research Methodology

This dissertation documents activities in the scientific discipline of IS research, which is part of the field of applied sciences and aims to "understand and improve the ways people create value with information" (Nunamaker & Briggs, 2011, p. 2). In this context, the research conducted for this dissertation is design-oriented, following ADR and consortium research as the basic methodological paradigms. Further research techniques used for the dissertation were literature review, case study, expert interview, focus group, plenary discussion, opinion survey, and reference modeling. The following subsections give a brief outline of each research method and technique applied.

4.1.1 Design Science Research (DSR)

DSR is a research method aiming at two goals: (1) to produce useful results that are of practical relevance, and (2) to advance the body of scientific knowledge. While practical relevance can be achieved by designing solutions to practical problems in the IS domain (Hevner et al., 2004; March & Smith, 1995; Walls, Widmeyer, & El Sawy, 1992), scientific relevance can be ensured by a rigorous research design that follows a consistent methodology (Winter, 2008). Typical design artifacts (i.e. results) of DSR are constructs, models, methods, and instantiations (March & Smith, 1995). To produce these artifacts, a number of authors have proposed certain principles and guidelines (Baskerville, Pries-Heje, & Venable, 2009; Hevner, 2007; March & Smith, 1995). One of the most recognized research designs was introduced by Peffers, Tuunanen, Rothenberger, and Chatterjee (2007) comprising six steps: (1) problem identification, (2) definition of design objectives, (3) artifact design, (4) artifact demonstration, (5) artifact evaluation, and (6) communication of results.

4.1.2 Action Design Research (ADR)

ADR is a combination of DSR and AR. In AR, the researcher aims at advancing the body of scientific knowledge by taking an active role and inducing change in the organization under analysis. AR consists of five steps: (1) diagnosis, (2) action planning, (3) action taking, (4) evaluating, and (5) specifying learning (Baskerville & Wood-Harper, 1996; Susman & Evered, 1978).

ADR was introduced in 2011 as “a research method for generating prescriptive design knowledge through building and evaluating ensemble information technology (IT) artifacts in an organizational setting” (Sein, Henfridsson, Purao, Rossi, & Lindgren, 2011, p. 40). The method is based on the observation that DSR artifacts always reflect the organizational context in which they are produced. Consequently, ADR is an integrative approach characterized by frequent changes of the practical and the scientific perspective in order to design useful (i.e. practically applicable) and rigorous (i.e. based on the use of scientific methods) artifacts. ADR consists of four stages: (1) Problem Formulation, (2) Building, Intervention, and Evaluation, (3) Reflection and Learning, and (4) Formalization of Learning (Sein et al., 2011).

4.1.3 Consortium Research

Consortium research is defined as “a multilateral form of collaborative research in which practitioners grant researchers access to their knowledge, collaborate in the specification of solutions, test artifacts in their business environments, and finance the research activities” (Österle & Otto, 2010, p. 283). In this sense, consortium research explicates the guidelines of ADR for a specific multilateral and longitudinal research setting characterized by close researcher-practitioner interactions. As an organizational design for engaged research (Back, Krogh, & Enkel, 2007), it proposes an iterative approach involving multiple partner companies in four phases: (1) analysis, (2) design, (3) evaluation, and (4) diffusion (Österle & Otto, 2010).

4.1.4 Research Techniques Applied

In addition to applying ADR in the context of consortium research, the author of the dissertation used several research techniques to design the key artifact. These techniques are outlined in *Table 4-1* and explained in the subsequent paragraphs.

Table 4-1: Research techniques applied

Tech-niques applied	Problem formulation (PF)	Building, intervention, and evaluation (BIE)	Reflection and learning (RaL)	Formalization of learning (FoL)
Literature review	Analysis of the scientific body of knowledge and the state of the art in practice	Construction of the artifact based on scientific body of knowledge and the state of the art	Theorizing about design decisions and alternative solution designs	Documentation of the body of knowledge; Presentation of knowledge gaps
Case study research	Situational inquiry and problem identification	Examination of situational solution designs and artifacts; Application and instantiation of the artifact (situational design) and evaluation	-	Documentation of case studies
Expert interview	Situational inquiry and problem identification	Analysis and evaluation of (emerging) solution designs and the artifact	-	Documentation of (emerging) solution designs and the artifact
Focus group	Exploration and confirmation of problems	Review of emerging solutions; Discussion or confirmation of design decisions and the artifact; Evaluation of the artifact	-	Presentation of design decisions and the artifact
Plenary discussion	Review and confirmation of problem analysis and requirements	Review and confirmation of design decisions and different versions of the artifact	-	Presentation of design decisions and the artifact
Opinion survey	Confirmation of problems and requirements	Evaluation of the artifact	-	-
Reference modeling	Documentation of the requirements	Construction of the artifact; Evaluation of the artifact	Reflection of design decisions against "guidelines for orderly modeling"	-

Literature Review

"A review of prior, relevant literature is an essential feature of any academic project. An effective review creates a firm foundation for advancing knowledge" (Webster & Watson, 2002, xii). The literature review aims at identifying and analyzing relevant contributions, defining knowledge gaps inducing further investigations, and providing a sound basis for the research activities to be conducted (Levy & Ellis, 2006; Vom Brocke et al., 2015).

Case Study Research

The case study is a frequently used instrument in qualitative empirical research. It is a suitable method if the subject to be studied is rather complex – as is the case with large enterprises, for example – and requires including its context in the analysis (Eisenhardt, 1989; Eisenhardt & Graebner, 2007; Yin, 2014). To conduct a case study, several research techniques need to be combined, such as document analysis, expert interviews, or observations. Case studies can be categorized by their purpose (i.e. exploratory, explanatory, or confirmatory case study), the number of cases considered (i.e. single-case or multiple-case study), and the degree to which the researcher is involved in the study (i.e. observational or participatory case study). In contrast to a single-case study, a multiple-case study allows replication of results from one case in other, comparable cases. Multiple-case studies typically result in "more robust, generalizable, and testable theory than single-case research" (Eisenhardt & Graebner, 2007, p. 27). Theoretical saturation is achieved when the researcher has reached a point in the research process where including any further cases would not yield any additional insights (Glaser, 1965). Participatory case studies are a specific form of AR; in addition to observing the scene, the researcher takes an active role by intervening in, and thereby directly influencing, the observed process (Susman & Evered, 1978).

Expert Interview

The expert interview is another widely used instrument in qualitative empirical research. Expert interviews are conducted with individuals who are regarded as subject matter experts in the field to be studied (Meuser & Nagel, 2009). Compared with other empirical methods, expert interviews allow a relatively fast accumulation of knowledge (Mieg & Brunner, 2004).

Focus Group

Unlike the expert interview, which is typically conducted with a single, or sometimes two or three subject matter experts, a focus group comprises a larger number of people (typically between five and 15 subject matter experts). The focus group is an interactive setting guided by a moderator, who asks the participants to share their opinions and experiences on a predefined topic, with the aim of explicating implicit knowledge and gathering in-depth expert feedback (Morgan, 1996). In the context of DSR, focus groups allow the identification and discussion of practical problems, as well as the design and evaluation of artifacts (Tremblay, Hevner, & Berndt, 2010).

Plenary Discussion

Similarly to a focus group, a plenary discussion is a moderated exchange of views and opinions within a group of peers on a predefined topic (Morgan, 1996). The topic is typically defined and introduced (e.g. in the form of an impulse presentation) prior to the discussion. In contrast to a focus group, a plenary discussion can comprise a large number of participants in the plenum to pass on practitioners' knowledge to academic researchers (Österle & Otto, 2010). A plenary discussion allows both exploration of practical problems and reviewing potential solutions.

Opinion Survey

An opinion survey is a method for collecting data in a quantitative empirical study. The data can be collected by means of (semi-)structured questionnaires (online or offline) or telephone interviews, for example. For data evaluation, methods of statistical analysis can be applied in order to discover common patterns across the answers of the participants, with the objective of reaching a certain level of generalization of the results gained from a sample to a larger entity (Gable, 1994; Pinsonneault & Kraemer, 1993; Siau & Rossi, 2011).

Reference Modeling

Reference modeling comprises a number of methods for designing reference models[26] (Fettke & Loos, 2007). Most methods propose an incremental design process, typically covering the following steps: (1) problem definition, (2) reference model design, (3) reference model evaluation, and (4) reference model application (cf. Becker, Algermissen, Delfmann, & Knackstedt, 2002; Becker & Delfmann, 2007). For supporting the design activities in accordance with scientific standards, several authors have developed "guidelines for orderly modeling"[27] (Becker et al., 1995; Schütte, 1998; Schütte & Rotthowe, 1998).

---

[26] see *Section 2.4.1* for an introduction to and a definition of reference models

[27] In German: "Grundsätze ordnungsgemäßer Modellierung". According to Becker, Rosemann, and Schütte (1995), reference model design has to follow six general requirements: correctness, relevance, cost effectiveness, clarity, comparability, and systematic structure (author's translation).

4.2 Context of the Research Conducted

The artifact to be constructed is a capability reference model, which supports enterprises in managing data as a strategic resource. Generally speaking, a reference model specifies the generally valid elements of a system and serves as an orientation for designing company-specific models (Fettke & Loos, 2004; Vom Brocke, 2007). The artifact is a practice-oriented conceptual model representing a selected phenomenon of interest and serving a certain purpose. By providing a common understanding of the subject, the artifact enables communication between different stakeholder groups, while supporting the design, implementation, and maintenance of a solution (Wand & Weber, 2002). To address these research goals, the author of the dissertation followed the methodology of ADR and consortium research in the context of the CC CDQ.

The CC CDQ is a consortium research program that was launched in 2006. Since then, it has comprised over 30 (mostly large, global) enterprises, as well as researchers from three high-profile universities (the University of St. Gallen, TU Dortmund University, and the University of Lausanne). Consortium research, being a multilateral and longitudinal form of DSR, acknowledges that there is a large body of knowledge both in the scientific and in the practitioners' domain (Österle & Otto, 2010). It establishes a formal collaboration between researchers and practitioners from various enterprises (Back et al., 2007), who agree on a joint research agenda, meet in regular workshops, and jointly conduct case studies. More than 60 workshops and more than ten doctoral dissertation projects have contributed to a constant accumulation of knowledge and experiences on data management in the CC CDQ. This body of knowledge significantly influenced the design of the artifact of this dissertation.

The research activities conducted by the CC CDQ since its establishment in 2006 can be divided into three phases (Legner et al., 2020), which reflect the changes in how data is viewed and what role it plays in companies (see *Table 4-2*):

(1) *ontology phase:* identifying the constituents of quality-oriented data management and creating a shared understanding of the subject,

(2) *capability-building phase:* assessing data management maturity and building up the capabilities required for effective quality-oriented data management, and

(3) *reorientation phase:* addressing the requirements of strategic data management.

Table 4-2: Research phases of the CC CDQ

No.	Name	Period	Research questions	Key artifacts
1	Ontology	2006 / 2007	What is data management? What are the key constituents of data management?	Initial version of the reference model (Framework for CDQM).
2	Capability-building	2008 - 2014	How can necessary capabilities for data management be built up? How can data management maturity be assessed?	Refinement of the reference model and its design areas via methods, tools, and guidelines. Initial version of the maturity model (Maturity Model for EDQM).
3	Reorientation	Since 2015	What are the key constituents of strategic data management? How must the artifacts developed so far be changed to meet the requirements of strategic data management?	Reorientation and revision of the reference model (DXM) and its design areas via methods, tools, and guidelines (*focus of this dissertation*). Modification and extension of the maturity assessment model.

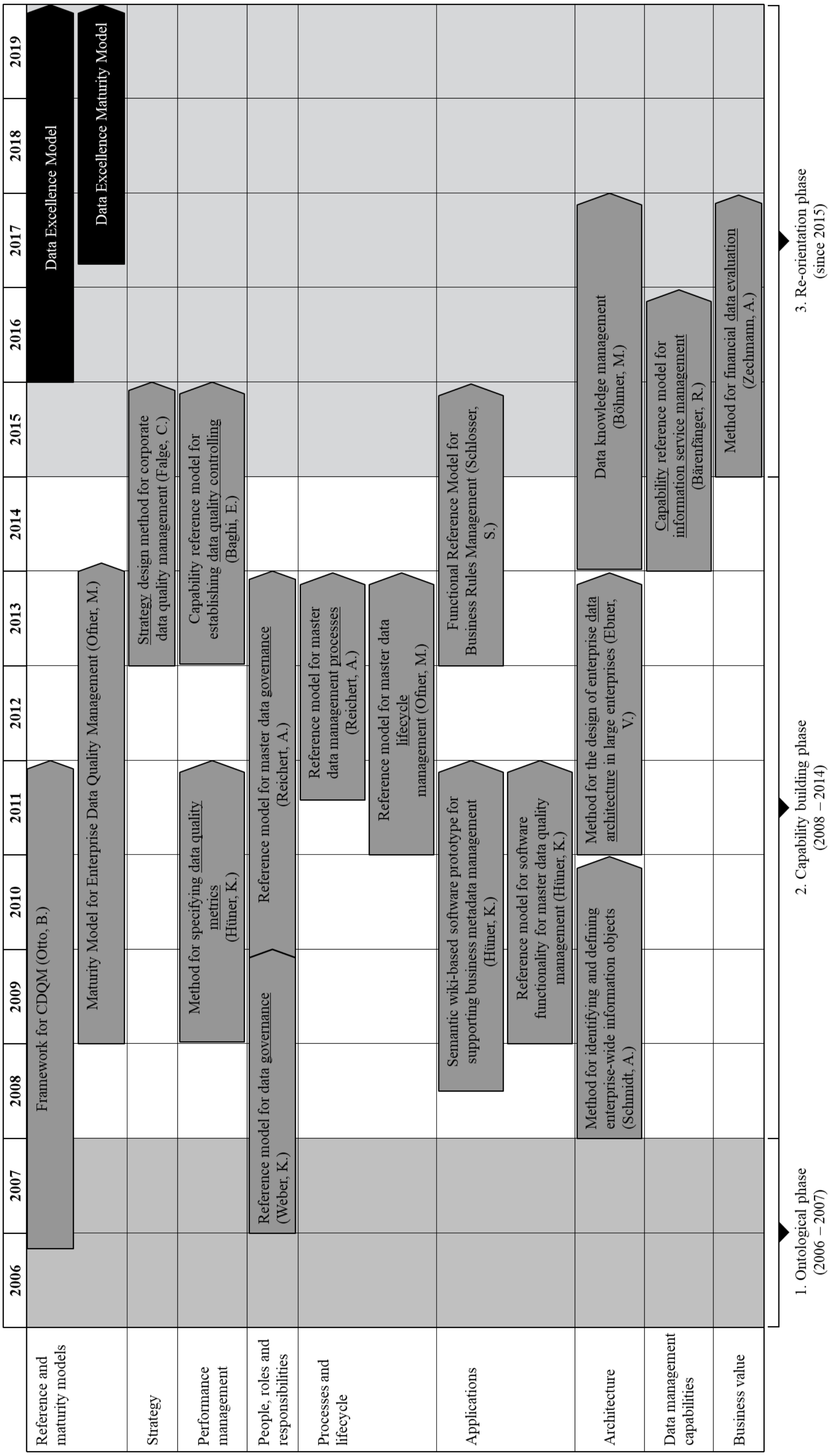


Figure 4-1: Research phases of the CC CDQ (based on (Legner et al., 2020)

Figure 4-1 presents an overview of the research activities across the three research phases of the CC CDQ, which are outlined in the following subsections. While the design activities towards the capability reference model for strategic data management took place in phase three, the first two phases (prior to the research activities of the author) are also presented in the following as the CC CDQ body of knowledge evolved along all phases and provided the context of the reorientation towards strategic data management.

4.2.1 Phase 1: Ontology (for Quality-oriented Data Management)

A group of data management experts from practice and academia founded the CC CDQ in 2006. Given the low maturity and significant challenges of the user companies with data management, the consortium research project's first two years focused on identifying common problems in enterprise-wide data management and developing a shared understanding of the subject. To do so, the consortium organized four focus groups and five plenary discussions, with the aim to define the requirements and objectives of enterprise-wide data quality management (DQM).

The research activities in this early phase of the program aimed at finding answers to fundamental questions:

- What are the key elements of quality-oriented data management?
- What is considered as part of quality-oriented data management?
- What is the rationale for addressing quality-oriented data management? What are the issues if this is not addressed?

To answer these questions, it was decided to develop a reference model for DQM. The companies of the CC CDQ had been asking for such an aid, which would help them define the key elements and the scope of DQM, and which they could use as an orientation for defining or adjusting their own, company-specific approach. Furthermore, a reference model would help them instruct their employees and communicate the importance of data management to the various stakeholder groups in their companies. The reference model design process, which was based on business engineering (Österle & Winter, 2003) as the conceptual foundation, resulted in the Framework for CDQM with its six design areas (see *Subsection 3.2.1*).

4.2.2 Phase 2: Capability-building (for Quality-oriented Data Management)

Over time, more and more companies started using and instantiating the Framework for CDQM to design, implement, and communicate their data management projects. With the companies' experience in handling the model growing, the research activities of the CC CDQ were shifted to increasingly address aspects of implementation:

- How can necessary capabilities for implementing quality-oriented data management be built up?
- How can quality-oriented data management maturity be assessed?

Based on the feedback received from the companies using the Framework for CDQM, the research activities in phase two focused on further specifying the model's six design areas. This was considered an important step in order to offer companies methodological guidance in building up data management capabilities. One of the first research activities of phase two was the definition of roles for data management, resulting in a reference model for data governance (Weber et al., 2009b; Wende, 2007). These roles and their respective responsibilities were then further specified (Otto & Reichert, 2010) and complemented by MDM processes (Reichert, Otto, & Österle, 2013). To support companies in data controlling, the researchers developed data quality metrics (Hüner, Schierning, Otto, & Österle, 2011; Otto, Ebner, & Hüner, 2010; Otto, Hüner, & Österle, 2009) and a capability reference model for data quality controlling (Baghi, Otto, & Österle, 2013). Furthermore, the researchers developed a reference model for data management applications (Otto, Hüner, & Österle, 2012) and methods for designing the data architecture (Baghi, Schlosser, Ebner, Otto, & Österle, 2014; Schmidt & Otto, 2008).

In parallel, the consortium established a second strand of research in order to address the member companies' need for monitoring the progress of data management and benchmarking with other companies. It was decided to take a closer look at the concrete practices required for enterprise-wide DQM, with the aim to determine the level of maturity a company has reached in data management. The activities for designing and evaluating a maturity model, which covered a period of five years, included several iterations and intensive collaborations with practitioners. These activities finally resulted in the Maturity Model for EDQM published in 2009 and 2013 for the scientific community (Hüner et al., 2009; Ofner, Otto et al., 2013) and in 2011 and 2016 for the practitioners' community (EFQM, 2011, EFQM, 2016) (see *Subsection 3.2.2*).

4.2.3 Phase 3: Reorientation (Towards Strategic Data Management)

In 2012, the CC CDQ started to discuss the increasing strategic relevance of data (see research activities 0-1 to 0-5 in *Table 4-3*). However, it was not until 2015 that the consortium realized that the notion of data being a strategic resource would have a significant impact on data management. In that year, the CC CDQ decided to fundamentally revise the Framework for CDQM, with the goal to provide companies with a reference model that supports them in the process of transforming into data-driven enterprises (research activity 1-1 in *Table 4-3*). The research activities in phase three of the competence center have therefore focused on analyzing the data management requirements of such an enterprise. This required to review the fundamental questions regarding quality-oriented data management – as specified in phases one and two – in a broader context:

- What should be (further) considered as part of strategic data management?
- What is the (extended) rationale for addressing strategic data management?
- How can necessary capabilities for implementing strategic data management be built up?
- How can strategic data management maturity be assessed?

As phase three is where the CC CDQ is still operating at the moment, the fundamental questions raised above have provided the setting and motivation of the research conducted for this dissertation.

4.3 Research Activities

The design of the capability reference model for the strategic data management followed the ADR methodology, which by definition is subdivided into four stages: (1) Problem Formulation (PF); (2) Building, Intervention, and Evaluation (BIE); (3) Reflection and Learning (RaL); and (4) Formalization of Learning (FoL) (Sein et al., 2011). During the ADR process, the author of the dissertation used the research techniques introduced in *Subsection 4.1.4*.

Figure 4-2 presents an overview of the ADR process. Each of the four process stages is detailed in the *Subsections 4.3.1* to *4.3.4*. *Subsection 4.3.5* concludes this section by listing and detailing the research activities conducted.

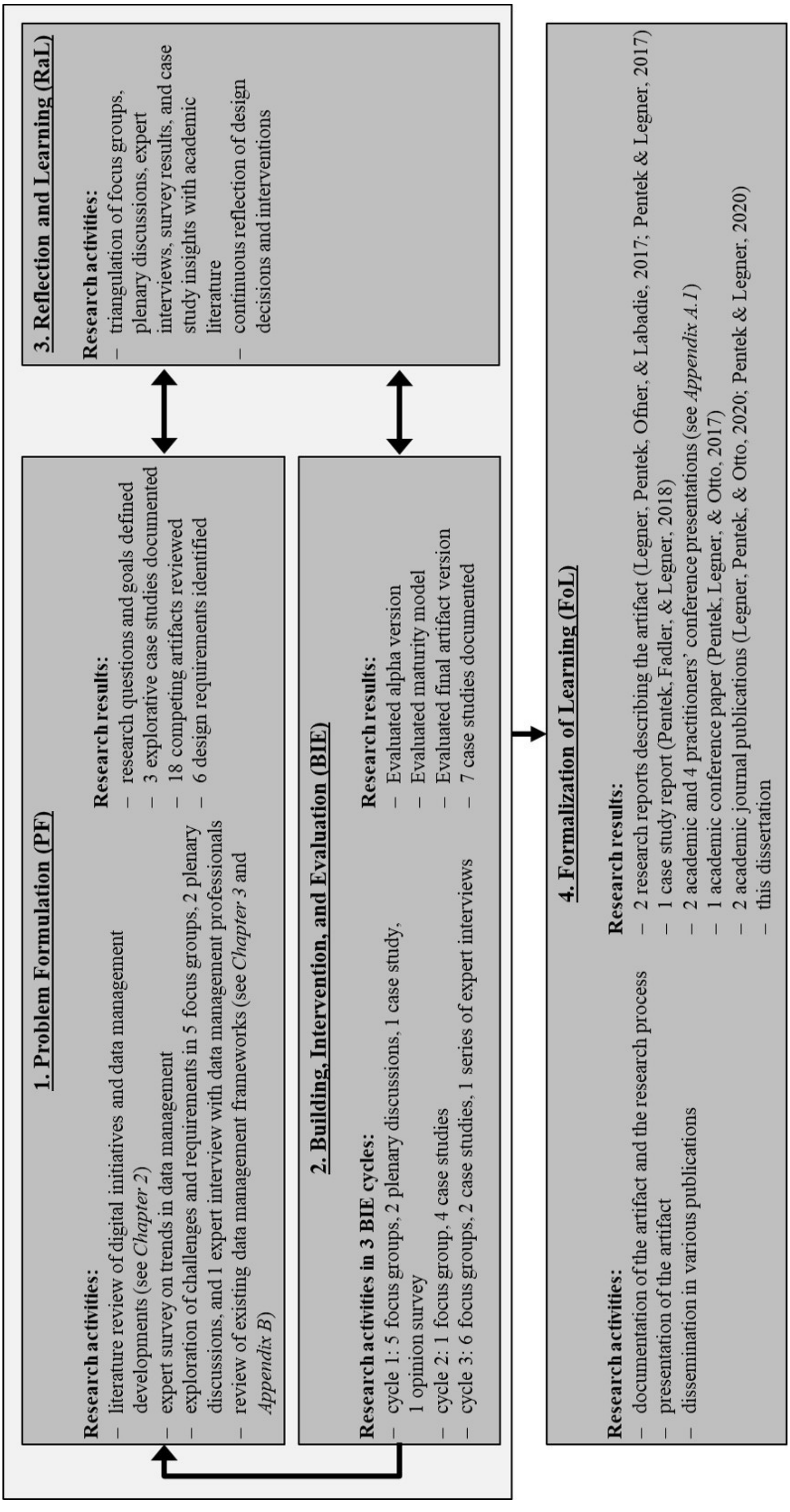


Figure 4-2: ADR process overview (based on Sein et al., 2011)

4.3.1 Problem Formulation (PF)

To explore how digital transformation and data-driven decision-making impact data management, the author first conducted a comprehensive literature review (see *Chapter 2*) including (amongst others) publications of the CC CDQ on digital initiatives and data management (Bärenfänger, Leveling, & Otto, 2016; Bärenfänger & Otto, 2015) and data management frameworks (see *Chapter 3*). After that, the author conducted an expert survey comprising 23 data management professionals from 19 companies from various industries (research activity 1-4). The author presented the findings from the literature review and the expert survey to three focus groups (research activities 1-3, 1-5, and 1-8) and in two plenary discussions (research activities 1-2 and 1-6) during consortium workshops. The discussion of the findings helped identify crucial practical problems of data management and defining design requirements to be met by a reference model for strategic data management. The author then documented and structured the results of the discussions and triangulated them with academic literature. Later in this stage, the author conducted three exploratory case studies to identify and examine rationales of data management other than data quality (research activity 2-6) and review approaches towards data management for data-based decision making and data-enabled products (research activity 3-2). The requirements on data management for such "offensive" data usage scenarios were discussed by another focus group (research activity 3-1). Finally, the author organized one more focus group specifying the requirements to be met by a maturity model for strategic data management (research activity 2-4)[28].

4.3.2 Building, Intervention, and Evaluation (BIE)

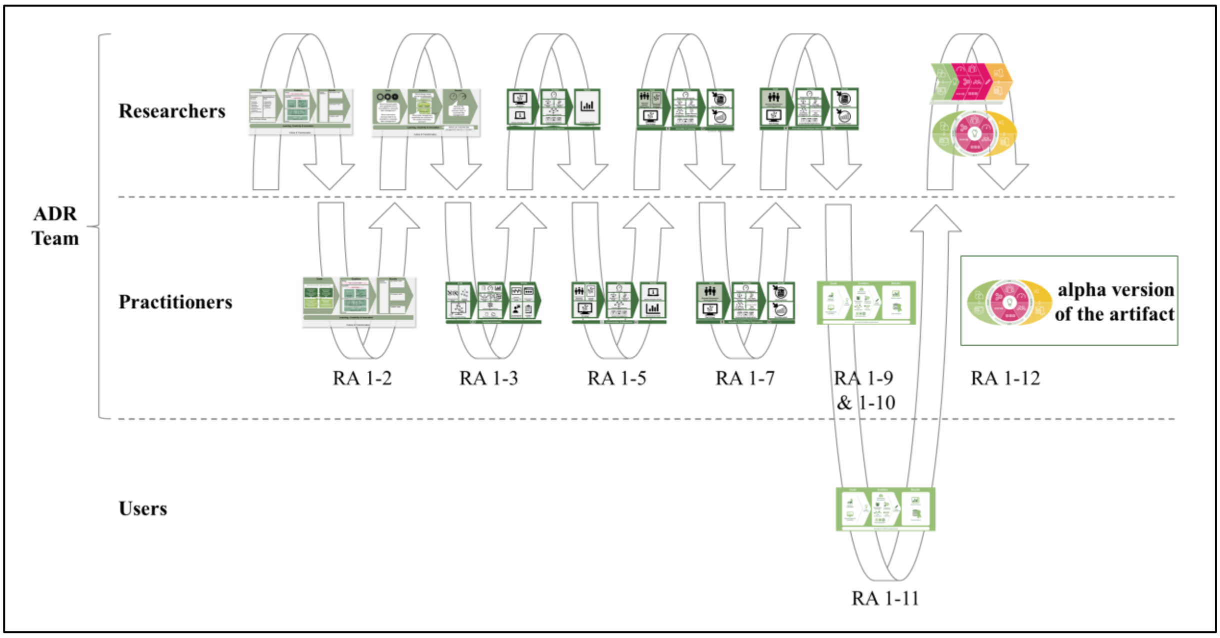
BIE took place in three cycles: the first BIE cycle (lasting from November 2015 to March 2017) resulted in an alpha version of the reference model defining twelve design areas as key constituents of strategic data management; the second cycle (April 2017 to March 2018) generated a maturity model detailing the capabilities of each design area; and the third cycle (April 2018 to January 2020) resulted in the final version of the capability reference model for strategic data management.

During each BIE cycle, the author of the dissertation presented working versions of the reference model to practitioners for being discussed and refined in plenary discussions and focus groups taking place during consortium workshops. Each session lasted about

[28] Although the development of a maturity model for strategic data management is not subject of this dissertation, the discussion of these requirements entailed content-related needs, which influenced the design areas and capabilities of the DXM as the domain model of the maturity model to be developed.

two hours, was moderated by the author of this dissertation, and was observed by a second researcher. The results of these discussions were thoroughly documented by the author as a basis for further adjustments of the reference model. Furthermore, plenary discussions, case studies, expert interviews, and an opinion survey complemented the research techniques applied during BIE.

The first BIE cycle started with a plenary discussion focusing on the structure of the reference model to be developed (research activity 1-2). After that, the author organized two focus groups for defining the relevant design areas of the model (research activities 1-3, 1-5, and 1-7). The author then organized another focus group for further specifying and naming the design areas and agreeing on a name for the reference model (research activity 1-9), leading to an initial version of the DXM. This version of the model was then formally evaluated by means of a questionnaire-based survey with data management experts (research activity 1-10). In response to the expert feedback, the author organized another focus group dealing with an improved graphical representation of the reference model to facilitate its use for communication purposes (research activity 1-12). In addition, the author conducted case study research (research activity 1-11) to further refine the reference model from a practical perspective, resulting in the alpha version of the DXM. *Figure 4-3* depicts the first BIE cycle with the respective versions of the DXM[29].



[29] *Appendix A.3* provides a more detailed overview how the visualization of the artifact evolved.

Figure 4-3: Building, intervention, and evaluation (BIE) cycle 1 (adapted from Sein et al., 2011, p. 42)

During the second BIE cycle, the alpha version of the reference model was instantiated by means of several case studies to further detail the design areas and capabilities of the DXM (research activities 2-1, 2-2, and 2-3). The author of this dissertation also revised the Maturity Model for EDQM (EFQM, 2011, EFQM, 2016; Ofner, Otto et al., 2013) to match the structure and design areas of the DXM. Based on an expert interview with data managers from a large transportation company (research activity 2-4), the author adjusted the maturity model and put it to the test in the course of a participatory case study (research activity 2-5), leading to slight changes of the reference model in turn. The maturity model resulting from this case study was then refined in a focus group discussion during a consortium research workshop (research activity 2-7) – again leading to further amendments of the DXM design areas and, especially, its success criteria and recommended practices. These criteria and practices were further refined in case studies applying the maturity assessment (research activity 2-8) leading to the final version of the DXM maturity assessment questionnaire. *Figure 4-4* depicts the second BIE cycle.

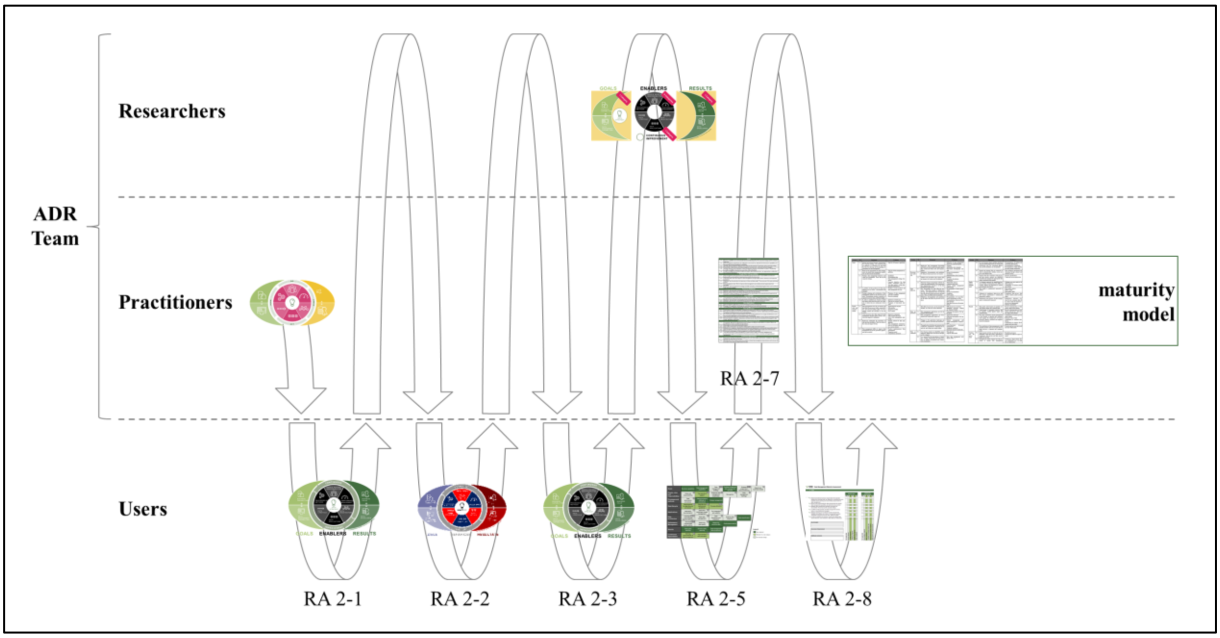


Figure 4-4: Building, intervention, and evaluation (BIE) cycle 2 (adapted from Sein et al., 2011, p. 42)

In the third BIE cycle, the alpha version of the reference model was instantiated and refined in two case studies (research activities 3-5 and 3-9). The reference model was further detailed through research activities focusing on data strategies (research activities 3-3, 3-6, 3-7, 3-8, and 3-9). Finally, two focus group discussions with data management researchers (research activities 3-10 and 3-12) resulted in the final version of the DXM, which includes a consistent terminological basis and the integration of previous and ongoing research results of the CC CDQ. *Figure 4-5* depicts the third BIE cycle.

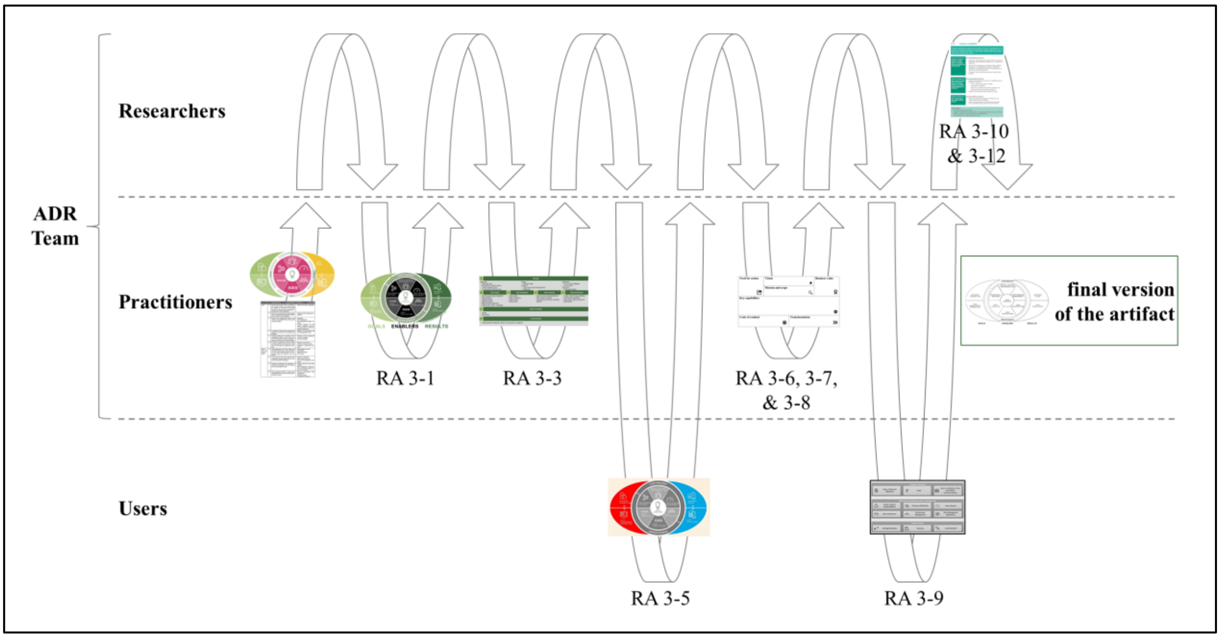


Figure 4-5: Building, intervention, and evaluation (BIE) cycle 3 (adapted from Sein et al., 2011, p. 42)

4.3.3 Reflection and Learning (RaL)

RaL activities were conducted in parallel to PF and BIE activities. This stage included the collection and documentation of practitioners' requirements, the collection of feedback given with regard to the reference model and the maturity model, the triangulation of results with academic literature, and the continuous reflection regarding the design and redesign of the reference model and maturity model.

4.3.4 Formalization of Learning (FoL)

The FoL stage generated formal descriptions of the reference model for being shared with and assessed by practitioners, and for dissemination across the academic community. Diffusion activities included the presentation of the reference model both at academic and practitioners' conferences (see list of presentations in *Appendix A.1*) and the publication of practitioner-oriented and academic contributions (research activities 1-13 and 3-11). The alpha version of the artifact was published as a research-in-progress paper (Pentek et al., 2017) and in a research report (Pentek & Legner, 2017). The final version was published in two academic journals (Legner et al., 2020; Pentek & Legner, 2020)[30]. Furthermore, the DXM was applied for communication purposes in a case study with a large enterprise from the fast-moving consumer goods (FMCG) industry (research activity 3-5, see *Subsection 5.5.2*) and in a data management education program (research activity 3-4, see *Appendix A.4*).

4.3.5 Research Activities Details

Table 4-3 details the research activities (including the design decisions made) in chronological order. It also refers to intermediate versions of the reference model, as depicted in *Appendix A.3*. In between these explicit research activities with practitioners, the author conducted additional model design and refinement activities on his own and with further CC CDQ researchers, as well as literature reviews for reflection of the research process and results. These activities are not listed in the table.

Table 4-3: Research activities conducted

ID	Date	Technique	ADR stage	Description
0-1	18.04.2012	Plenary discussion	Problem formulation (PF)	Discussion of the need for companies to 1) integrate external data (e.g. data from smart meters or e-mobility) and 2) be compliant with data management-related regulations.
0-2	22.06.2012	Focus group	PF	Discussion of the need for companies to 1) assess the impact of new technologies (e.g. big data analytics) on data management and 2) extend the scope towards data security management.
0-3	13.02.2013	Plenary discussion	PF	Discussion of the need for companies to calculate the profitability of data management.
0-4	09.10.2013	Focus group	PF	Discussion of the need for companies to have data architectures for big data scenarios at their disposal.
0-5	29.10.2014	Focus group	PF	Discussion of the need for companies to address data management requirements posed by digitalization and Industry 4.0.
1-1	04.11.2015	Focus group	PF	Agreement to revise the Framework for CDQM.

[30] *Appendix A.5* provides a full list of all publications of the author related to this dissertation.

ID	Date	Technique	ADR stage	Description
				Discussion of the need for companies to develop data management services and capabilities for the digital economy. Details in *Appendix A.2*
1-2	25.02.2016	Plenary discussion	PF, Building, intervention, and evaluation (BIE)	Discussion of the requirements to be met by a reference model for strategic data management (e.g. taking business criticality and business value of data into account; addressing compliance and data security aspects; extending the scope to all data types). Discussion of the concept of the continuous management cycle (i.e. PDCA cycle: Plan – Do – Check – Act) as the basic structure of the reference model. Design Decision 1 (Conceive of data management as a continuous management cycle) and Design Decision 6a (Demonstrate results): The discussion group confirmed the proposal of the research team to include 1) Goals, 2) Enablers, 3) Results, and 4) Learning, Creativity & Innovation in the reference model to be developed (see *Figure A-1* in *Appendix A.3*).
1-3	29.04.2016	Focus group	PF, BIE	Specification of data management requirements to be met by the reference model (e.g. taking business criticality and business value of data into account). Review of existing reference models and maturity models for data management. Discussion of relevant design areas of data management (e.g. change management not as a design area of its own; “Data Strategy” subsumed under “Goals”). Details in *Appendix A.2* Design Decision 3 (Explicate a data strategy) and Design Decision 4 (Develop data management capabilities through organizational and technological aspects): The focus group agreed to adopt the elements of the CDQM Framework for the reference model to be developed, i.e. Controlling, Organization & People, Processes & Methods, Corporate Data Architecture, and Corporate Data Applications as Enablers, and to subsume Data Strategy under the Goals category of the model (see *Figure A-2* in *Appendix A.3*).
1-4	05/2016 - 08/2016	Opinion survey, expert interview	PF	Identification of current trends in data management by means of semi-structured expert interviews, identifying business analytics, Industry 4.0, and ecosystem collaboration as relevant drivers of data

ID	Date	Technique	ADR stage	Description
				management. Companies are extending the scope of data management beyond master data, while considering compliance, security, privacy and risks as key priorities on top of data quality. See (Legner et al., 2017)
1-5	29.06.2016	Focus group	PF, BIE	Discussion of requirements to be met by the reference model (e.g. linking data management with business), and of the design areas and structure of the reference model. Details in *Appendix A.2* Design Decision 5 (Emphasize the data lifecycle) and Design Decision 6b (Demonstrate results in terms of data excellence and business value): The focus group decided to define Data Lifecycle as a separate design area in the Enablers section of the model, and not as part of Processes & Methods. Furthermore, the group agreed to distinguish between business impact and data impact in the Results section of the reference model (see *Figure A-3* in *Appendix A.3*).
1-6	14.09.2016	Plenary discussion	PF	Presentation and discussion of survey results (see 1-4).
1-7	14.09.2016	Focus group	BIE	Discussion of relevant design areas and the terminology for the design areas of the reference model (e.g. “Performance Management” instead of “Controlling”; “Data Excellence” as an umbrella term for data-related results of data management). Details in *Appendix A.2* Design decision 2 (Translate business capabilities into data management capabilities): The focus group decided to integrate the RBV and a capabilities perspective into the Goals section of the reference model, and to link data management capabilities to business capabilities (see *Figure A-4* in *Appendix A.3*).
1-8	10.11.2016	Focus group	PF	Identification of requirements to be met by data management (e.g. ensuring compliance with regulations such as GDPR; addressing the management of open data). Request to revise the CDQ Academy format based on the new reference model. Details in *Appendix A.2*
1-9	09.12.2016	Focus group	BIE	Discussion of design area names and contents (e.g. “People, Roles, and Responsibilities” instead of “Organization”; “Business Value” instead of “Business Impact”).

ID	Date	Technique	ADR stage	Description
				Agreement to name the model "Data Excellence Model". Details in *Appendix A.2*
1-10	09.12.2016	Opinion survey	BIE	Formal, questionnaire-based evaluation of the reference model (see *Figure A-5* in *Appendix A.3*). See *Subsection 7.2.1*
1-11	20.02. - 21.02.2017	Case study	BIE	Early alpha version of the DXM (see *Figure A-5* in *Appendix A.3*) tested by a large enterprise from the pharma industry for communicating with data management stakeholders and developing a data strategy. See Bayer case study 1 in *Subsection 5.4.1*
1-12	22.02.2017	Focus group	Formalization of learning (FoL), BIE	Presentation of case study results (see 1-11). Discussion of layout options (see *Figure A-8* in *Appendix A.3*) and agreement on a graphical representation of the Data Excellence Model. Details in *Appendix A.2*
1-13	11/2016 – 03/2017	-	FoL	Preparation of DESRIST contribution and research report presenting the alpha version of the DXM.
2-1	19.04.2017	Case study	BIE	Instantiation of the alpha version of the DXM by a large enterprise from the manufacturing industry for sensor data management to prove the validity of the reference model for "new" data domains. See Schaeffler case study in *Subsection 5.4.4*
2-2	03/2017 - 05/2017	Case study	BIE	Instantiation of the alpha version of the DXM by a large enterprise from the transportation industry to develop a governance policy for MDM. See SBB case study 1 in *Subsection 5.4.3*
2-3	04/2017 - 12/2017	Case study	BIE	Instantiation of the alpha version of the DXM by a large enterprise from the pharma industry to develop a corporate directive for MDM. See Bayer case study 2 in *Subsection 5.4.2*
2-4	04.09.2017	Expert interview	PF	Identification of the requirements to be met by a revised maturity model by means of expert interviews (e.g. ensuring compatibility with previous version; extending the scope of the questionnaire; discussion of the reference model's beta version). Details in *Appendix A.2*
2-5	09/2017 - 12/2017	Case study	BIE	Design of a maturity model matching the DXM in collaboration with a large enterprise from the transportation industry. See SBB case study 2 (*Subsection 5.5.1*)

ID	Date	Technique	ADR stage	Description
2-6	11/2017	Case study	PF	Exploration of Data Ethics and Business Value Management at a large enterprise from the software industry. See SAP case studies in *Subsections 5.3.2* and *5.3.3*
2-7	08.12.2017	Focus group	BIE	Presentation of the results of the case study for designing the maturity model (research activity 2-5). Refinement of maturity model elements and questionnaire. Details in *Appendix A.2* and in *Appendix C.1.*
2-8	01/2018 – 03/2018	Case studies	BIE	Application of the maturity model based on the DXM at ABB, Bayer, Beiersdorf, Schaeffler, Shell, Swarovski, and Zespri.
3-1	25.04.2018	Focus group	PF, BIE	Discussion of the differences between managing big data and managing master data and company-internal data (e.g. defining additional roles, such as data scientist; processes such as data science use case identification, and architectures for data lakes). Details in *Appendix A.2*
3-2	27.04.2018	Case study	PF	Exploration of data management in a data-driven enterprise. See PMI case study in *Subsection 5.3.1*
3-3	22.06.2018	Focus group	BIE	Discussion of data (management) strategies and their elements to support business capabilities and create business value. See description of Data Strategy (*Subsection 6.5.3*) and details in the *Appendix A.2*
3-4	Since 09/2018	Focus group	FoL	Application of the DXM in a data management education program. See CDQ Academy description in *Appendix A.4*
3-5	10/2018	Case study	BIE, FoL	Application of the DXM by a large enterprise from the FMCG industry for communicating MDM activities across all stakeholder groups. See tesa case study in *Subsection 5.5.2*
3-6	09.05.2019	Focus group	PF, BIE	Review of current status, drivers, and scope of data strategies in CC member companies See description of Data Strategy (*Subsection 6.5.3*) and details in *Appendix A.2*
3-7	26.06.2019	Focus group	BIE	Review of two exemplary data strategies (from Bosch and SBB) and discussion of the Data Strategy Canvas See description of Data Strategy (*Subsection 6.5.3*) and details in *Appendix A.2*

ID	Date	Technique	ADR stage	Description
3-8	08/2019 – 10/2019	Expert interviews	BIE	Series of 23 expert interviews on current status and drivers of data strategies, and analysis of the key elements and the relationship with other corporate strategies and initiatives See description of Data Strategy (*Subsection 6.5.3*) and details in *Appendix A.2*
3-9	09/2019	Case study	BIE	Application of the DXM by a large enterprise from the automotive industry for the purpose of developing a data strategy. See Bosch case study in *Subsection 5.4.5*
3-10	16.10.2019	Focus group	BIE	Discussion of success criteria and recommended actions per design area. Details in *Appendix A.2*
3-11	10/2017 – 12/2019	-	FoL	Preparation of JAIS and HMD publications presenting the final DXM version.
3-12	20.01.2020	Focus group	BIE	Discussion of success criteria and recommended actions per design area. Details in *Appendix A.2*

5. Case Studies

In the process of observing, designing, instantiating, and evaluating the central artifact of this dissertation, case study research played a prominent role. Generally speaking, case study research aims at investigating "a contemporary phenomenon […] in depth and within its real-world context" (Yin, 2014, p. 16). It is considered suitable if the phenomenon of interest cannot, or should not, be isolated from its environment, and if it is still largely unexplored at the time the research is conducted (Benbasat, Goldstein, & Mead, 1987; Yin, 2014). The author used three types of case studies for his dissertation project: (1) *exploratory case studies*, for a better understanding of the phenomenon of interest during the problem formulation stage, (2) *participatory case studies*, for designing, instantiating, evaluating, and redesigning the artifact, and (3) *confirmatory case studies*, for evaluating the artifact's utility. The insights gained from these multi-case research activities complement the findings from the literature review conducted by the author (see *Chapter 2* and *Chapter 3*) and add a practical perspective on the phenomenon of interest.

This chapter starts with a short description of the data collection process and the instruments used therefore (*Section 5.1*), before the process of selecting the cases is briefly outlined (*Section 5.2*). It then introduces three exploratory case studies, which were documented during the problem formulation stage of the research process: one conducted with PMI (*Subsection 5.3.1*) and two conducted with SAP (*Subsections 5.3.2* and *5.3.3*). The experiences and results gained from these three case studies had a substantial impact on the design of the reference model as they illustrate good practices from comparably mature data-driven enterprises.

Afterwards, this chapter present seven case studies illustrating the model design activities and demonstrating possible model instantiations, giving proof of the model's general applicability and practical utility. The case studies reflect two typical scenarios for applying the reference model in practice: (1) translating the abstract design knowledge of the reference model into a concrete situational design, and (2) applying the reference model as abstract design knowledge for communication, education, maturity assessment, and benchmarking purposes. Accordingly, this chapter at first presents five case studies instantiating a situational design at Bayer, SBB, Schaeffler, and Bosch (*Section 5.4*), before it outlines two case studies conducted with SBB and tesa illustrating the model's application as abstract situational knowledge (*Section 5.5*). This chapter closes with a summary and discussion of the presented case studies (*Section 5.6*).

5.1 Data Collection

The case study research activities conducted by the author of this dissertation aimed at gaining a better understanding of the phenomenon of interest (i.e. Problem Formulation), designing the artifact in close collaboration with practitioners (i.e. Building and Intervention), and evaluating and redesigning the artifact (i.e. Evaluation and Re-Building). In total, the author conducted and analyzed ten case studies[31].

The companies the case studies were conducted with are members of the CC CDQ community. For data collection, a semi-structured expert interview was the main method used. Each interview (either face to face or over telephone) was documented in the form of hand-written notes, which were then transcribed and sent to the interviewee for review and approval. In addition, data was gathered from company-internal sources (e.g. presentations, strategy papers, guidelines, standard operating procedures) and from public sources (e.g. websites, annual reports, press reports). This data was used for triangulation of the case study findings (Benbasat et al., 1987; Senger & Österle, 2004; Yin, 2014). Data collection took place between February 2017 and September 2019. *Appendix A.4* outlines the interviews conducted and the additional sources used for documenting the case studies. All case study documentations were sent to the respective company representatives for review and approval.

5.2 Case Selection

Each case study documents a project or initiative related to data management in a given company setting. The selection of cases for the dissertation project followed the recommendations of Eisenhardt (1989) and Eisenhardt and Graebner (2007) regarding theoretical sampling. The following criteria were decisive for the selection of cases:

(1) **Reflection of the different phases of data management:** For being selected, a project or initiative had to cover either the "traditional", quality-oriented approach of data management, or the advanced, strategic approach of data management, or both.

(2) **Focus on the three core management aspects of the artifact to be developed (i.e. Goals, Enablers, Results):** The cases were chosen in a way that each aspect was covered by at least two cases. According to Eisenhardt, 1989), this multi-

[31] In addition, another application of the DXM for structuring a data management training concept was examined. This example is presented in *Appendix A.4*.

case approach allows researchers to replicate findings per category, leading to high reliability of results.

(3) **Generalizability of results:** The cases represent large, multinational companies of different size (with regard to the number of employees and annual revenue) and from different industries.

Table 5-1 classifies the selected cases based on these three criteria. Furthermore, the table indicates at which stage or stages of the research process the case study was conducted. The case studies are presented in *Sections 5.3*, *5.4* and *5.5*. *Sections 5.6* and *7.1* discuss the case studies' contribution to the design and evaluation of the capability reference model.

Table 5-1: Case study classification

	1a) Quality-oriented data management	**1b) Strategic data management**	**2a) Focus on Goals**	**2b) Focus on Enablers**	**2c) Focus on Results**	**3) Large, multinational company**	**4a) Problem formulation**	**4b) Building and intervention**	**4c) Evaluation**	**4d) Reflection and learning**	**4d) Formalization of learning**
1. PMI	X	X	X	X	X	X	X	-	-	-	-
2. SAP 1	(x)	-	-	X	-	X	X	-	-	-	-
3. SAP 2	-	X	-	-	X	X	X	-	-	-	-
4. Bayer 1	X	(x)	X	-	(x)	X	-	X	X	-	-
5. Bayer 2	X	-	-	X	-	X	-	X	X	-	-
6. SBB 1	X	-	-	X	-	X	-	-	X	-	-
7. Schaeffler	X	X	X	X	X	X	-	(x)	X	-	-
8. Bosch	X	X	X	X	X	X	-	-	X	-	-
9. SBB 2	X	(x)	X	X	X	X	X	X	X	-	-
10. tesa	X	-	-	X	-	X	-	-	X	-	X
X: fully addressed; (x): partially addressed; -: not addressed											

5.3 Case Studies Exploring Strategic Data Management

This section presents the three exploratory case studies conducted in relation with this dissertation project during the problem formulation stage. Following the recommendations for case study reports given by Senger and Österle (2004), each case study description introduces the company, depicts the initial situation, describes the activities performed during the project and the solutions resulting thereof, and reflects the lessons learned from the case.

5.3.1 PMI: Pursuing a Data-driven Business Transformation

The PMI case presented in this subsection is a revised excerpt of the CC CDQ case study "PMI's Journey Towards a Data-driven Enterprise" published by Pentek, Fadler, and Legner (2018).

Company Profile

Philip Morris International Inc. (PMI) is a leading company in the global market of combustible tobacco products. It owns six of the world's top-15 cigarette brands (among them Marlboro). PMI's tobacco products are sold to over 150 million consumers in over 180 countries. PMI operates and owns 46 production facilities in 32 countries (see *Table 5-2* for details).

Table 5-2: Philip Morris International company overview

Philip Morris International (PMI)	
Founded	1900 Until its spin-off in 2008, PMI was an operating company of Altria Group.
Headquarters	Operational headquarters: Lausanne, Switzerland Corporate headquarters: New York, US
Industry	Cigarettes and smoke-free products
Revenues	78.1 billion USD (2017)
Earnings before interest and taxes (EBIT)	11.5 billion USD (2017)
Number of employees	80,600 (2017)

Initial Situation

As a response to changing consumer preferences and the growing trend towards health consciousness across all industrialized countries, PMI decided to make a fundamental change regarding its business model and product portfolio. In developing and selling alternative products that do not require burning or combustion, the company is now pioneering the way to a smoke-free future. Over the last ten years, PMI invested about 4.5 billion USD in the research and development of smoke-free products, employing about 400 dedicated scientists, engineers, and technicians in this process. An important milestone was the launch of IQOS, PMI's smoke-free tobacco heating system, in Japan and Italy in 2014. Two years later, the company started its business transformation by stating its ambition to ultimately give up the combustible tobacco products business in favor of the development and distribution of alternative, smoke-free products.

This strategic shift implies that PMI is moving away from its historically grown core business, which has been to produce a single type of product distributed predominantly over B2B sales channels. Among other things, PMI's new business model and product portfolio implies manufacturing and selling electronic devices for consuming tobacco, developing and establishing new manufacturing processes, and developing and establishing forms of direct interaction with consumers by applying a B2C business model.

In PMI's strategic shift, data plays a key role. The company's Chief Executive Officer (CEO), André Calantzopoulos, formed the notion of PMI transforming into a science company, with its consumers at its core. The move towards a B2C business model and the creation of a digital business function requires PMI to better understand and respond to consumers' changing needs. PMI's senior management realized the strategic importance of data already during the phase of planning the company's transformation.

In 2016, Jacek Olczak, PMI's Chief Financial Officer (CFO) at the time, mandated a small project team of three people to develop a rough concept and structure for PMI's new, data-driven approach supporting the company's new vision of a "smoke-free future". In three months, the team did an analysis of the company's status quo and specified what needed to be done in terms of data governance by looking at best practices from other companies and industries.

For PMI, data represents both a challenge and an opportunity: a challenge, because PMI's entire data landscape proved unable to facilitate data-driven business decisions in the context of the company's transformation; and an opportunity, because in the wake of the megatrend of digitalization, new technologies and approaches have become available, allowing efficient and successful data governance, generating actionable, data-based insights, and facilitating data-driven business decisions, all of which can help accelerate PMI's business transformation.

PMI has historically collected data for the purpose of manufacturing and selling combustible tobacco products. Business functions were individually generating and gathering data. Except for a few incomplete initiatives, data governance was a low priority at PMI, and was not considered at an enterprise-wide level. In the early 2000s, a first initiative established a governance model for finance master data and operations master data. However, due to the company permanently growing, flaws in the data governance organization occurred, and the data structures that were set up became obsolete. Hence, PMI's data landscape was characterized by a number of shortcomings, such as the existence of data silos, excessive data duplication, multiple sources of truth, and inconsistent data definitions. In addition, there was no central data governance unit with a corporate-wide perspective on data, as each business function had its own data experts and data stewards.

In 2016, which was the baseline year of assessing the status quo, PMI ran a central data management application – a customized SAP instance – containing limited master data sets, such as product codes, organizational codes, or financial information. Much of the data was redundant, as it was available in other systems as well. As a result, there were a lot of duplicate data records containing different values. Another problem at that time was that a small team of experts from the Information Systems (IS) function conducted some form of data science, but without having a clear role and skillset descriptions at their disposal, and without following a defined demand-gathering process. Consequently, the activities of this group were not given a lot of attention.

To make this business transformation a success story, three major requirements have been defined: (1) A transformation like this will not be possible on the basis of an ungoverned data landscape. (2) The new business and operating model, leading to an ever-increasing amount of data, requires a data governance framework. (3) Newly available technologies need to be fully understood and extensively applied to allow PMI to leverage big data and make data-driven decisions.

Solution

PMI's transformation into an enterprise with the consumer at its core is very closely linked to the establishment of a new function: Enterprise Data & Analytics (EAD). The EAD team pursues the goal of disrupting and transforming the way people at PMI think about and use data. It does so by

- ensuring that data is managed as a corporate asset, based on the concept of data ownership, the use of clear data definitions and useful data models, and the specification of proper data processes and business rules to establish a SSOT;
- integrating data science and data analytics into decision-making processes;
- replacing decision-making based on gut feeling and individual experience by fact-based (i.e. data-based) decision-making; and
- building an internal, company-wide team of disruptors equipped with a data-driven mindset, skills, and competencies.

In financial terms, EAD's goal was to generate five times more business value than EAD cost. To demonstrate transparency regarding the business value of EAD's activities, the EAD team selects and evaluates data-driven use cases.

To implement the vision of a data-driven enterprise, PMI decided to establish two major pillars for the EAD organization:

(1) Data Governance, ensuring a trusted and consistent data foundation by defining and controlling data ownership, data policies, data-related principles, and data-related standards across all functions; and

(2) Data Science, to experiment with data and to design and reuse algorithms to generate business value by providing actionable insights.

PMI regarded this dual strategy as crucial for effectively addressing data-related issues, breaking data silos, and turning data into business opportunities. Each of the two major pillars was subdivided into a number of building blocks: Data Governance was subdivided into Data Architecture, Data Management, Data Quality, and Data Development, while Data Science was split up into Data Science and Data Science Enablement. For each of these six building blocks, a team of specialized staff was established. The teams were supported by the Program Management Office (PMO) and a Data Protection team. Data protection activities took place outside the EAD organization, as they were conducted by PMI's IS and legal teams. *Figure 5-1* depicts the building blocks of the two major pillars constituting PMI's EAD organization in June 2018.

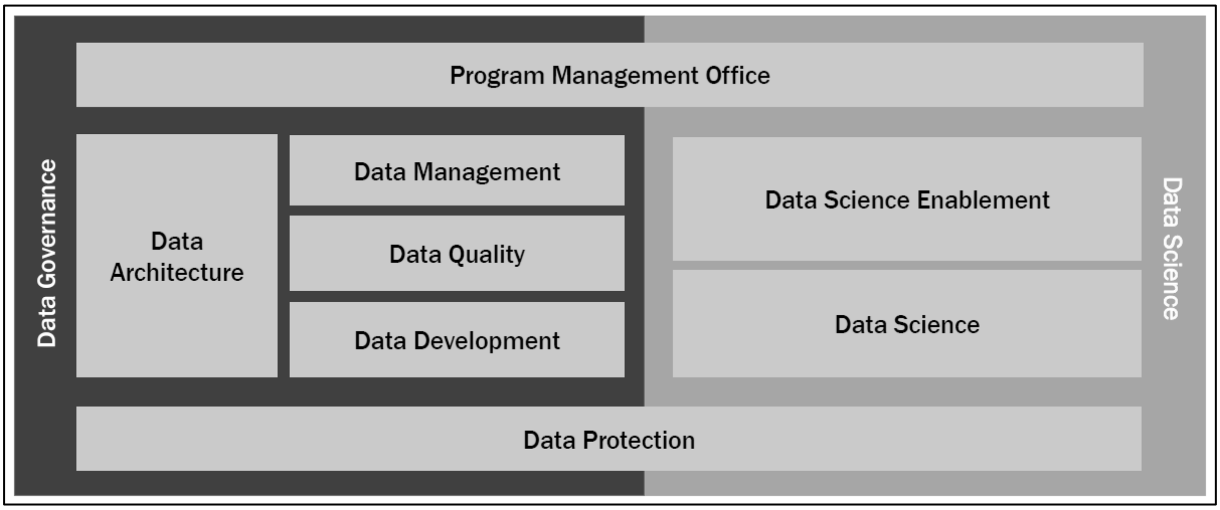


Figure 5-1: PMI's EAD organization (in June 2018) (from Pentek et al., 2018, p. 13)

Phase 1

The organizational setup of EAD took place in two phases. During the first phase (from the beginning of 2017 until June 2018), the priority was on establishing data governance and delivering first results from data science. In this phase, EAD focused on four building blocks, namely Data Management, Data Architecture, Data Science, and the Program Management Office (PMO), each of which reached a considerable level of maturity by June 2018.

The **Data Management** team established enterprise-wide data governance structures, processes, and standards. As data experts were distributed across the entire organization, a technical solution needed to be established which allowed these experts to collaborate efficiently. To achieve this goal, the team selected Collibra as its platform for data governance. The platform was implemented at a very early stage in the process of EAD's transformation (in early 2017). Collibra provides data definitions, data structures, and data sources, as well as data-related rules and responsibilities to support data governance. It is constantly updated and populated every time a business function is in contact with EAD. Collibra comprises two main modules: (1) the business glossary, providing high-level information to business users who are looking for specific information (e.g. definitions of a certain term, a certain data entity/attribute, or a certain key performance indicator (KPI)); and (2) the data dictionary, providing technical information, such as data profiles or data lineage information, but also information on data systems and data models. Both modules are connected to each other, so that there is a bridge from business to IT allowing users to navigate from a business term directly to the corresponding data entity. From an application perspective, the vision regarding MDM is to have one central hub which a number of systems have access to. For this purpose, the Data Management team selected SAP MDG, which is now connected to two main target systems: PMI's ERP system and the commercial analytics application. From an operational perspective, the Data Management team is also responsible for master data maintenance. In August 2016, the first of PMI's four Shared Service Centers, which are responsible for data maintenance in the six geographic regions PMI has specified for its business activities, was integrated into the EAD organization. Through the implementation of SAP MDG, each Shared Service Center offers a single workflow for initiating maintenance activities. These activities are conducted semi-automatically, allowing the service team to direct its attention to data quality and data validation as well as to functional support. By summer 2018, the process of integrating the first Shared Service Center into the EAD organization was completed. The remaining Shared Service Centers still operate separately at the moment, but will also be integrated into EAD on a mid-term basis. A major outcome of the activities of the Data Management team has been the development and establishment of a maturity assessment model. This model allows assessing data maturity in all of PMI's business functions, thereby creating transparency with regard to the status quo and providing a baseline for tracking EAD's progress. Based on the maturity assessments conducted, the team could identify and document 39 owners of 692 data entities. Moreover, the team specified six guidelines for managing data assets.

To facilitate the capture of the data definition and data ownership, the Data Management team collaborates closely with the **Data Architecture** team. This team is responsible for the conceptual and the logical data models, as well as for data integration. It makes sure that the models are interconnected across the different domains in order to ultimately map the overall data landscape of the organization. The Data Architecture team maintains Collibra, PMI's data governance platform creating transparency regarding all data available, and how the data is interrelated. For PMI's business transformation, the Data Architecture team takes into account different perspectives on data: an enterprise-wide view, an internal, consumer-centric view, and a business process view. By July 2018, after 18 months of operation, the Data Architecture team had already documented over 690 data entities and over 4,840 data attributes in Collibra. The team has developed, and now maintains, 27 conceptual models and 54 logical models, referring to several domains (including product data, device data, customer data, consumer data, and business partner data). For example, one model the team is in charge of developing is an enterprise-wide data model for devices used across the entire company; this data model will be applicable across all functions and for different applications (repair center software, e-commerce desktop solutions etc.).

The **Data Science** team is responsible for generating actionable, data-based insights across PMI. The team first focused on the organizational setup of four Data Science Labs (in Lausanne, Amsterdam, Krakow, and Tokyo), the establishment of basic procedures, and the implementation of the technical infrastructure: PMI's *Data Ocean*, the exclusive place to perform professional Data Science exploration. Beyond this, the Data Ocean is the authoritative platform for analytics regarding data originating from cross-functional or external sources, and for storage of external core data (i.e. data not generated by PMI's own systems). Activation of the Data Ocean started in 2017 by (1) establishing the foundation (i.e. activation of the platform in AWS plus activation of Data Science Workbench, the Integration Data Layer, and a governance framework), (2) delivering early prototypes to validate the ecosystems of the technologies used (which are mainly open-source technologies), and (3) providing prove of the business value of the platform for implementing PMI's data and analytics strategy. It then continued in 2018 with the stabilization, scaling, and automation of the platform. Based on PMI's business priorities, the Data Ocean in June 2018 contained more than 40 business critical datasets and five complex data products. As a result of the newly defined demand-gathering process and the use case delivery model developed by the Data Science team, 25 use cases had been approved by mid-2018. While one of these use cases has been industrialized outside the Data Ocean meanwhile, seven are in preparation for being industrialized, and eight are in the delivery phase. In total, Data Science has delivered use cases representing a high double-digit number, in million USD, of potential costs savings or additional revenues.

The **Project Management Office** (PMO) coordinates all of EAD's activities and orchestrates change management activities within EAD and across the entire company. The PMO has three main objectives: (1) team alignment, (2) marketing and communication, and (3) education. The PMO team has developed and continuously updated EAD's strategic roadmap summarizing the value proposition of EAD's products, short-term and long-term objectives, and corresponding activities. For communication and educational purposes, the team established a platform for raising awareness of the importance of data and building up a data-related knowledge base.

Starting out with six people in January 2017, the EAD team has grown to over 100 people by the end of 2017, and to over 140 people by June 2018 (round about 45 in Data Science, 74 in Data Management, 13 in Data Architecture, and four in the PMO).

Phase 2

The second phase (starting June 2018 and scheduled to be completed by the end of 2019) has been designed to focus on substantiating the remaining three building blocks, namely Data Quality, Data Development, and Data Science Enabling.

Activities regarding **Data Quality** – covering the definition of data quality rules, audits, tools, interventions, and improvements – have deliberately been placed in the second phase of the organizational setup of EAD, as data quality presupposes a certain level of maturity of the Data Management and the Data Architecture building blocks.

Activities regarding **Data Development** – aiming at optimizing data-related processes and reporting processes in PMI's business functions and key business initiatives – were scheduled to start subsequently, as data development presupposes a certain level of data quality.

Activities regarding **Data Science Enablement** – aiming at developing capabilities to gather, evaluate and prioritize demand, and to plan and monitor capacities – also were scheduled to start subsequently and be conducted in parallel with Data Development activities.

Lessons Learned

PMI's case can be considered an example of a forward-looking and comprehensive transformation from a traditional business into a data-driven enterprise. The business transformation at PMI changes the perception and role of data, as the analysis of data and the generation of actionable, data-based insights across all business functions becomes a key success factor. In line with DalleMule & Davenport, 2017), the case illustrates an example of a company practicing a data strategy comprising both defensive and offensive elements. At PMI, this is achieved by the two core pillars of the EAD organization: Data Governance and Data Science. PMI has realized that it has to create a data foundation, which is achieved by Data Governance as a defensive element of the company's data strategy, *and* that it is necessary to generate actionable, data-driven insights and gain value from data, which is accomplished by Data Science being an offensive element of the data strategy (see *Table 5-3*).

Table 5-3: PMI's dual data strategy

	Defensive aspects	Offensive aspects
Key objectives	Create a data foundation through Data Governance	Generate business value by deriving actionable, data-driven insights through Data Science
Core activities	Phase I (Q1 17 – Q2 18): Establish enterprise-wide Data Governance, covering Data Management (i.e. processes, principles, standards) and Data Architecture Phase II (Q3 18 – Q4 19): Establish Data Quality and Data Development	Phase I (Q1 17 – Q2 18): Establish Data Science capabilities; experiment, and provide first insights; define demand-gathering process and use case delivery model. Phase II (Q3 18 – Q4 19): Establish Data Science Enablement
Enabling architecture	Single source of truth (SSOT) using SAP MDG for master data and Collibra for data definitions	Multiple versions of truth (MVOT) using the Data Ocean

By establishing a data foundation to break data silos, and by taking measures to avoid the most prevalent shortcomings (e.g. duplicate data, multiple sources of truth, or inconsistent data definitions), the approach pursued by PMI's EAD function clearly shows the characteristics of a defensive data strategy. Core activities in this approach refer to the optimization of data extraction, data standardization, data storage, and data access. Offensive elements of PMI's data strategy include optimization of data analytics, data modeling, data visualization, data transformation, and data enrichment, all of which aims at generating actionable, data-based insights. In contrast to a defensive data strategy, an offensive strategy puts great emphasis on flexibility. To achieve flexibility and facilitate MVOT, PMI established the Data Ocean platform, offering a reusable integration data layer and allowing systematic classification of data by means of a metadata catalog.

In addition to gaining the insight that pursuing a dual data strategy is beneficial, four further key lessons learned can be identified from the PMI case:

(1) The radical transformation of its business strategy (towards a smoke-free future) and business model (from B2B to B2C sales channels) has required PMI to rethink how to use and manage data. To acquire involvement and support from top

management, data management needs to link its activities to the business, create transparency and promote the value contribution of data to the business.

(2) A well-structured data organization strengthens the development of data management capabilities and can act as a driver of data-driven business transformation.

(3) Each data-driven business transformation requires a cultural change and a shift in the mindsets of all stakeholders involved. Therefore, employees need to be “taught” data and convinced to believe in the strategic importance of data to fully exploit the potential of data-driven insights. As many companies that have been around for a long time are characterized by a low level of data literacy, building data-related capabilities requires some time and a certain level of willingness to pursue such a dramatic shift.

(4) Each data-driven business transformation should be seen as a journey spanning several years. As a consequence, a successful data strategy approach needs to be adaptive to allow for continuous improvement.

5.3.2 SAP 1: Promoting Data Citizenship

Company Profile

SAP SE is a leading company for enterprise application software and software-related services. In terms of market capitalization, SAP is the world’s third largest software company. The company serves more than 425,000 customers in over 180 countries (SAP, 2019) (see *Table 5-4* for details).

Table 5-4: SAP company overview

SAP	
Founded	1972
Headquarters	Walldorf, Germany
Industry	Enterprise application software
Revenues	24.7 billion EUR (2018)
Operating profit	7.2 billion EUR (2018)
Number of employees	96,498 (2018)

Initial Situation

In 2009, SAP initiated a strategic program to improve the quality of its data. The program defined a data governance model, including roles such as Data Manager, Data Leader, or Friend of Data. For each of these roles, communication and training activities took place, leading to a significant improvement of data quality. Nevertheless, the data management team identified further potential for data quality improvement, as the program did not address the total of employees using and/or editing data.

Solution

From 2013 on, SAP pushed the program further to target the most important group when it comes to data quality: data users. The core idea of the initiative was that everyone should feel responsible for the data they use. The data management team started to collect examples of unfavorable behavior and action on the part of data users, leading to poor data quality for the company as a whole. Based on this collection, the team developed a list of principles of good data citizenship. For instance, one principle requests from data users to “search for data before you create duplicates”, while another demands that data users “do not create dummy data – if you are unsure, ask for help”. The general tone of the program activities was constructive. “We do not want to punish; we want to raise awareness. If there is a user community whose processes have a negative effect on data quality, we address their business lead, explain what effects poor data quality has for other users, and work together to address the root cause,” the program manager from the data management team pointed out.

Today, a dedicated member of the data management team is responsible for coordinating all marketing and change management activities of the program. In the first two years, the program was advertised to address a broad audience – internally via articles posted on the enterprise portal, posters hung up on the walls of offices and hallways, or laptop stickers given to employees; externally by communicating the program to customers and analysts or raising awareness for it at industry conferences. Especially the positive feedback coming from outside the company generated rewards to the data management team and to data users. After the initial awareness-generating phase, marketing activities were shifted to more personal communication with selected target groups, using instruments such as roadshows, training sessions, informal lunch breaks, or videos made with business leaders emphasizing the importance of data quality that were put on the enterprise portal. These activities not only helped raise awareness of and provide education for data management, but also contributed to a further improvement of data quality.

Lessons Learned

In the data-driven enterprise, practically each employee uses data one way or the other. She or he thereby becomes a “data citizen”. The term “citizenship” describes the state of being a member of a particular community that has certain privileges and duties (Cambridge University Press, 2014). Thus, a data citizen has both rights and responsibilities with regard to data. Above all, each data citizen is granted the right of having convenient access to all corporate data they need for performing their daily routines. Responsibilities mainly refer to sharing data, maintaining data, and using data in accordance with legal provisions and company internal requirements. The concept of data citizenship has two implications:

(1) The “communities of practice” dealing with data are not limited to data specialists anymore (such as Data Managers or Data Analysts), but include all data citizens. They form a large group across the entire enterprise, with each data citizen having specific knowledge, skills, and experience.

(2) Making sure that data citizens act according to their rights and responsibilities is a key task for data managers. This requires active change management, including the empowerment of every data citizen to become data literate. Apart from understanding the general importance and value of data, data literacy includes the ability to properly read, analyze, and utilize data. Consequently, data managers must define distinct roles for data citizens to assume, taking their individual data competencies into account. Finally, data managers need to make sure these competencies are continuously developed further and maintained.

5.3.3 SAP 2: Establishing Data Ethics

Initial Situation

In 2015, SAP’s data management department was approached by their colleagues from the marketing department asking for support in their activities regarding big data (mainly customer data) coming from multiple sources. Around the same time, SAP shifted its focus towards the protection and ethical use of customers’ personal data, following extensive discussions held in various political institutions of the EU regarding the need for a European data protection regulation[32], including the idea of imposing significant fines for non-compliant behavior.

Solution

To find an answer to the question as to how ethical use of customers' personal data can be ensured, SAP's data management department teamed up with the company's global risk and compliance department and data privacy officers. The group jointly developed a questionnaire, which was sent to relevant stakeholders to identify personal data, source systems, and the owners responsible for managing this data. Based on the answers from and interactions with these stakeholders, the group developed another questionnaire comprising a set of key questions to be answered by the owners of new, big data scenarios. That questionnaire included questions such as the following:

- Does the scenario comply with our brand strategy and corporate values?
- Is the scenario discriminatory (e.g. profiling by religious beliefs)?
- Could the scenario be regarded as too invasive?
- Is the scenario transparent with regard to what data is collected and for what purpose?

SAP then created a data ethics council comprising data leaders from across the company to collaborate with the owners of big data scenarios. Based on the questionnaire, the council today analyzes and discusses each new big data scenario and makes sure it complies with SAP's ethical standards.

Lessons Learned

Companies are increasingly affected by regulatory and data protection requirements – such as GDPR, which aims at giving the power back to data subjects by allowing them to define precisely what data collectors may do with their data and what not. These data subjects – such as customers providing both personal data and transactional data when they buy something online – are becoming increasingly sensitive about how their data is handled by companies. Since data is highly business critical, trust is essential – i.e. trust in the quality of data, the (internal and external) sources the data comes from, and in other functional departments or business partners (with regard to them using data as agreed upon).

---

[32] See *Subsection 2.2.5* for further information on the EU's General Data Protection Regulation (GDPR).

It should be pointed out here that simply adhering to regulations and laws is not sufficient to convince data subjects to reveal their personal data. Trust in the company collecting the data, together with its attitude towards data usage and processing, plays a key role here. Consequently, data managers should not only comply with regulatory requirements, but develop own, company-specific policies that go beyond what is legally demanded. This implies that data managers need to actively define how sensitive data must be handled, and that they need to establish mechanisms that ensure adherence to these standards. Finally, trust to adhere to jointly agreed data quality and data handling rules plays a key role in the willingness to share data with internal and external stakeholders.

5.4 Case Studies Instantiating a Situational Design

This section and the following section present seven case studies illustrating the model design activities and demonstrating model instantiations, giving proof of the model's general applicability and practical utility. The case studies reflect two typical scenarios for applying the reference model in practice: (1) translating the abstract design knowledge of the reference model into a concrete situational design, and (2) applying the reference model as abstract design knowledge for communication, education, maturity assessment, and benchmarking purposes. The following subsections present five case studies conducted with Bayer, SBB, Schaeffler, and Bosch instantiating a situational design.

5.4.1 Bayer 1: Developing a Master Data Management Strategy

Company Profile

Bayer AG was founded more than 150 years ago. Today, it is one of the leading chemical and pharmaceutical companies in the world. Its operations are managed in three divisions (Pharmaceuticals, Consumer Health, and Crop Science) and one business unit (Animal Health). In the self-care market, Bayer owns leading brands, such as Aspirin, Bepanthen, or Alka-Seltzer. The Bayer Group comprises 241 companies in 79 countries throughout the world (Bayer, 2018a, Bayer, 2018b) (see *Table 5-5* for details).

Table 5-5: Bayer company overview

Bayer	
Founded	1863
Headquarters	Leverkusen, Germany
Industry	Health care
Revenues	35.0 billion EUR (2017)
Earnings before interest and taxes (EBIT)	5.9 billion EUR
Number of employees	99,820 (2017)

Initial Situation

Bayer's Integrated Business Operations (IBO) is a global business service function responsible for most master data consuming back-office processes, such as order-to-cash or source-to-pay. In total, IBO covers six core processes. In addition to designing, governing, and conducting these processes and managing regional shared service centers, IBO hosts an MDM team, which is responsible for the management of enterprise reference data, business partner master data, bank and accounting master data, and material master data. In early 2017, IBO established a cross-functional master data governance organization including a Global Data Owner Community. Its activities were supported by a central enterprise data management group located in the IT department. Furthermore, IBO's main MDM processes – e.g. for maintaining global vendor master data or creating new customer master data – were defined.

In this situation, the global data owner community was faced with several external and internal requirements for MDM at Bayer. The enterprise needed to, and still needs to, comply with strict regulatory requirements impacting data management – such as the European Federation of Pharmaceutical Industries and Associations' (EFPIA) disclosure code[33], the European Medicines Agency's (EMA) request for identification of medicinal products (IDMP)[34], or GDPR. While the Global Data Owner Community was facing the challenge of motivating the usual master data stakeholders, Bayer's digitalization activities raised awareness of MDM among additional stakeholders posing further

[33] The EFPIA disclosure code requires pharmaceutical companies to disclose the transfers of value to healthcare professionals and healthcare organizations. To report the values, customers and the associated transactions need to be identified, matched, and aggregated across all business units, regions, and applications.

[34] IDMP requires pharmaceutical companies to report master data about medicinal products for human use (i.e. substance, product, and organization master data) in a standardized format to EMA.

data quality requirements. Or, as Jens Peter Henriksen, Global Process Owner for Business Partner Master Data at Bayer, explains: “We in master data were becoming victims of our own success. The organization and management were getting bored with the old stories about the importance of master data. There had not been any major catastrophes in recent years to remind people about the impact of bad master data. So, we needed a fresh perspective and a positive storyline to get the organization excited again.” As a consequence, IBO decided to develop a strategy and roadmap for MDM.

Solution

To develop the strategy, the author of this dissertation supported the Global Data Owner Community in designing, preparing, and moderating a two-day workshop taking place in February 2017 (see *research activity 1-11* in *Table 4-3*). An early alpha version of the DXM (see *Figure 5-2*) provided the structure for these activities. The workshop was focused on specifying the Goals section of the reference model for IBO by identifying relevant business capabilities, deriving the data management capabilities required, and developing a data strategy.

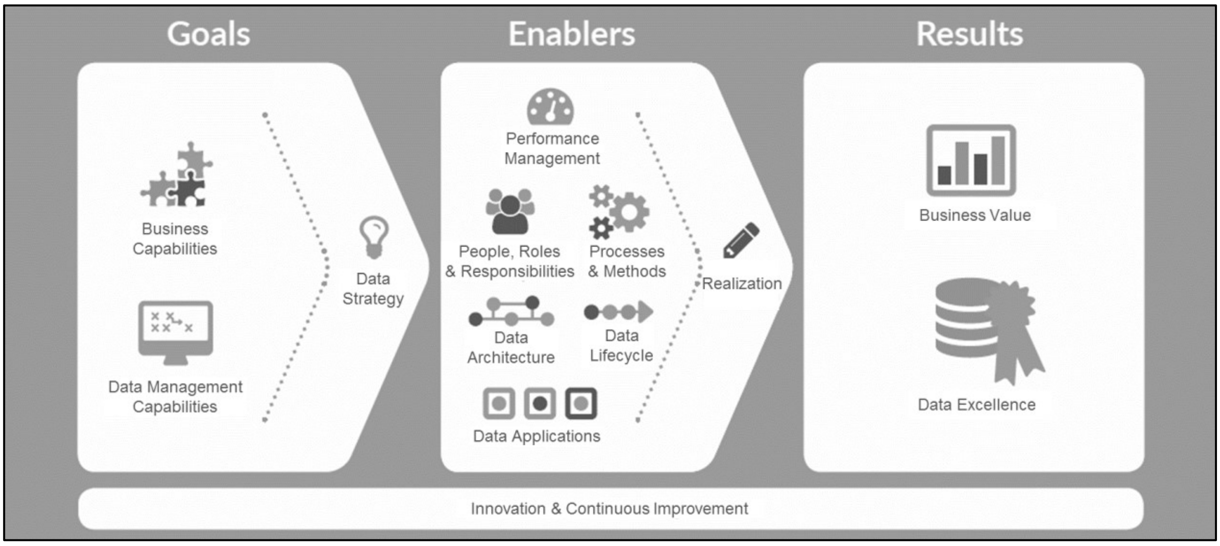


Figure 5-2: Early alpha version of Data Excellence Model

In total, 18 participants – data owners, IT representatives, and data consumers – engaged in the workshop (see *Appendix A.4* for participant information). After an introduction of the reference model, the participants were asked to split into six working groups, one for each of the six IBO core processes. Each group, then, discussed the goals of the respective core process, identified the business capabilities relevant for reaching these goals, and derived data management capabilities to support and enable the business capabilities (see *Figure 5-3* for the results gathered by the working group dealing with the reporting process).

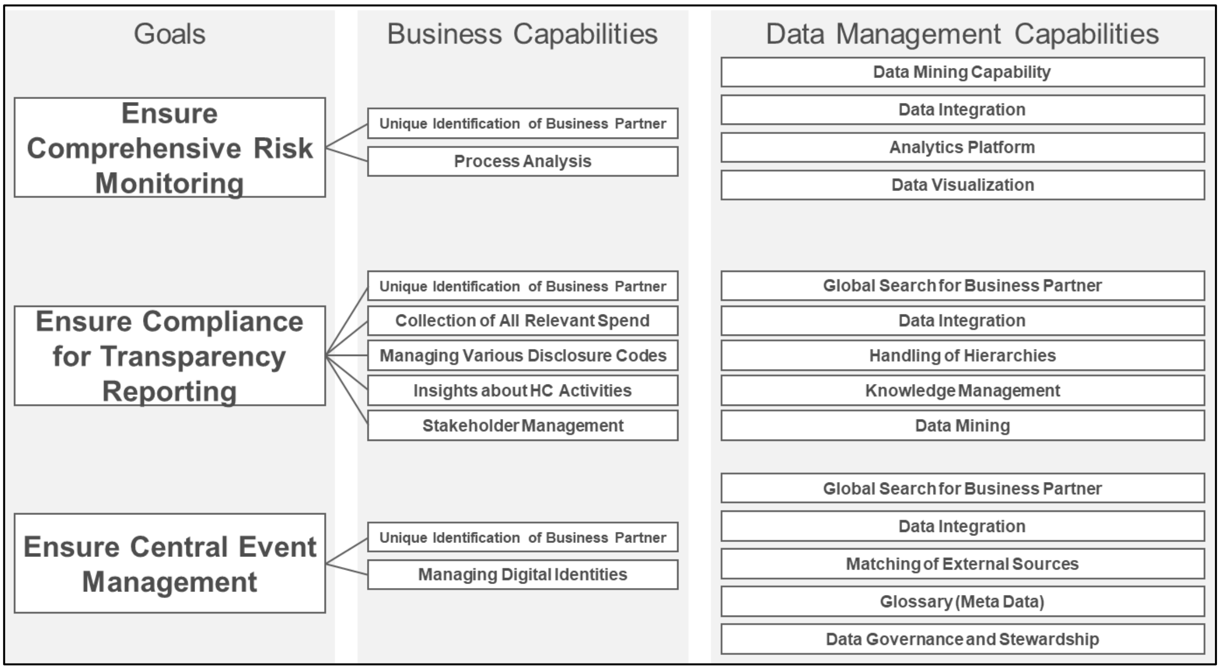


Figure 5-3: Goals, business capabilities, and data management capabilities of IBO reporting process

In another working group session, the participants were asked to assess the current status of each data management capability derived (choosing between "capability fully available", "capability partially available", and "capability not available"). Based on this assessment, during a moderated discussion, the participants prioritized and clustered the data management capabilities and defined four process-overarching focus areas for the MDM strategy (see *Figure 5-4*). Finally, during another moderated discussion, the participants defined a roadmap for implementing the strategy based on the data management capabilities and the strategic goals.

The resulting master data strategy has the vision of data becoming the "strong foundation for our transforming business". In order to reach the strategic goal of data excellence, Bayer's IBO function defined four focus areas for the company's master data strategy:

- *integration* through cross-functional data governance and harmonized data-related processes,

- *insights* based on well-described, standardized high-quality data,
- *compliance* addressing regulatory requirements, data security, and data integrity, and
- *agility* through a lean data management organization and efficient processes.

The strategic roadmap included two key activities: (a) redefining the master data governance organization, and (b) redefining global master data processes (see the Bayer 2 case study in *Subsection 5.4.2* for further details about these activities).

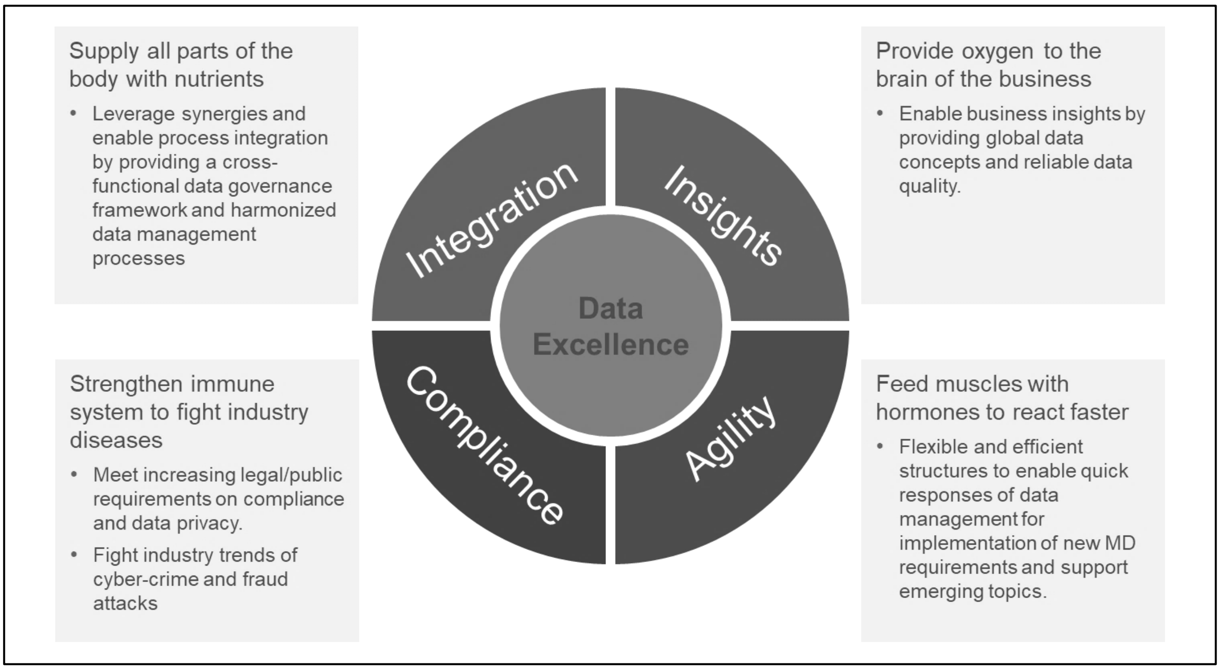


Figure 5-4: Strategic IBO master data management goal and focus areas

Lessons Learned

The workshop based on the alpha version of the DXM provided evidence of the general applicability and usefulness of the reference model for structuring and defining data management activities. Especially the Goals section – linking business capabilities, data management capabilities, and the data strategy – was regarded as very helpful, as the following quotes from a survey conducted among the participants indicate:

- "Great that we focus on business capabilities as a starting point. The outside-in view."

- "Challenging questions and exercises. The methodology helped to structure."
- "Good to focus strategy discussion on business value and future needs instead of 'only' traditional MDM talk on tools and governance models."

Overall, the workshop participants rated the reference model and the workshop structure positive (see *Figure 5-5*). Seven out of ten participants assessed the reference model's content as good or even excellent, while eight out of eleven said the workshop's concept and structure had been good or excellent.

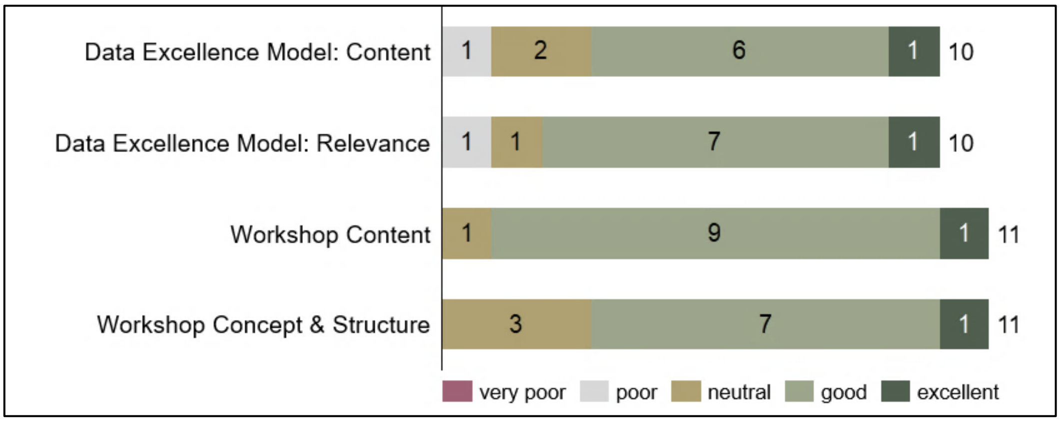


Figure 5-5: Participants' assessments of reference model and workshop

Furthermore, the workshop participants suggested explaining the capability view of data management in the model's Goals section, as reflections on the workshop had revealed the following observation: "The model looks simple but is complex to apply for many business users. 'Capabilities' is still an abstract term not yet commonly adopted. Careful explanation, strong moderation, and sufficient time is needed to get to the defining ('aha!') moments" (Henriksen, 2017, p. 16).

5.4.2 Bayer 2: Developing a Corporate Master Data Management Directive

This subsection documents the development of a corporate directive for MDM at Bayer. The activities described in the following took place after Bayer's IBO function developed a master data strategy with the help of an early version of the reference model (see previous *Subsection 5.4.1* for details on the strategy development and the company profile).

Initial Situation

In May 2013, Bayer issued its first corporate directive for the management of master data. This corporate directive, like any other corporate directive at Bayer, had to be revised after three years. Due to internal re-organization activities, however, Bayer's Master Data Governance Council postponed the revision by one year. In April 2017, the revision activities were initiated by the council (which consists of members from multiple business functions, such as BI, IBO, corporate supply chain, production, HR, and R&D). The author of this dissertation participated in these activities.

As a first activity, the members of the Master Data Governance Council provided feedback on the directive published in 2013. This survey revealed a number of shortcomings, the most prevalent of which were the following:

- Poor applicability: The document was perceived as too abstract and generic (as many stakeholders with divergent interests had been involved in the development process).
- Narrow scope: Having its focus on master data only, the directive neither covered transactional data nor metadata.
- Missing elements: The document did not address regulatory requirements impacting data management (e.g. IDMP, data privacy) and lacked any understanding of data being of significance for Bayer's digitalization activities.

Responding to these shortcomings, a revised version of the corporate directive was supposed to be providing a binding framework with principles and standards for MDM at Bayer. Furthermore, it had to be ensured that master data is available in an adequate quality and that regulatory and compliance requirements are met.

Solution

Bayer's Master Data Governance Council decided to apply the DXM as the guiding framework for developing the new version of the corporate MDM directive. Consequently, the directive introduces and outlines the DXM, and its core chapter is structured in accordance with the six design areas of the reference model's Enablers section (see *Figure 5-6*).

Table of contents

Figure 5-6: Table of contents of Bayer's corporate directive on master data management

The six Enablers design areas are specified by the directive as follows:

- **Processes and Methods** provide an overview of 39 major MDM processes at Bayer (see *Figure 5-7*). To reflect the logic of the DXM, this overview, which is based on Reichert's (2015) reference model for MDM processes, was enhanced by integrating the RBV (i.e. business capabilities and data management capabilities) and the concept of data excellence. As a result, Bayer's strategic, governance-related and operational MDM activities can be structured into nine process groups:
  - *MDM business alignment processes* ensure that the business capabilities are addressed by supporting data management capabilities, and that the targeted results of data management contribute to the business.
  - *MDM strategy processes* define the scope and objectives of data management and specify the roadmap for providing the data management capabilities required.

- o *People, roles and responsibilities processes* define the required MDM roles and boards, their interactions and responsibilities, and the necessary skills and training measures.
- o *Standards and guidelines processes* develop, review and update the standards, methods, and procedures required for MDM.
- o *Master data lifecycle management processes* detail the sourcing, creation, maintenance, transformation, release, archiving, and deletion activities for every data domain.
- o *Master data applications processes* identify functional requirements, design technical solutions, and initiate the implementation of these solutions.
- o *Master data architecture processes* model master data and its metadata.
- o *Performance management processes* establish a measurement system to monitor and control the progress and outcome of all MDM activities.
- o *MDM support processes* aim at training the roles and boards involved in MDM, and providing support to business, projects, and IT.

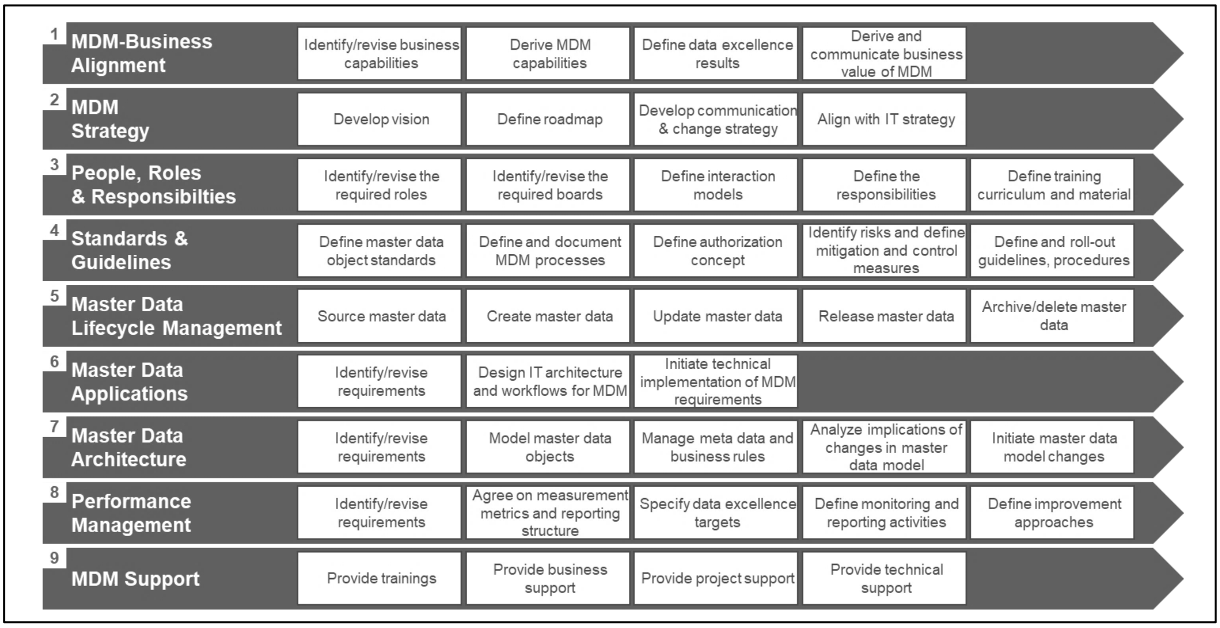


Figure 5-7: MDM processes at Bayer (based on Reichert et al., 2013)

- **Roles and Responsibilities** define roles and boards for conducting the above listed processes. The role model includes four dedicated MDM roles and three IT roles:
  - The *Global Process Owner – Master Data* specifies the MDM directive for her or his function and develops a functional master data strategy.
  - The *Functional Data Steward* implements the (functional) master data governance framework and defines standards.
  - The *Operational Data Steward* ensures harmonized execution and improvement of MDM processes.
  - The *Master Data Coordinator* implements changes and improvements on a local or regional level.
  - *IT Functional Leadership* ensures that Bayer's IT strategy is reflected by all master data activities.
  - *IT Design and Build* supports MDM in translating business requirements into compatible master data solutions.
  - *IT Execute* ensures operational excellence of applications and services.

  Furthermore, four boards are in charge of directing, governing, designing, and improving Bayer's MDM activities (see *Figure 5-8*):

  (1) The *Master Data Governance Council* aligns all MDM strategies on a corporate group level and ensures their compliance with Bayer's business strategy.

  (2) The *Master Data Governance Sub-Council* is responsible for the functional master data strategy and for ensuring its alignment with Bayer's business strategy.

  (3) The *Master Data Domain Board* is responsible for the master data demand management process on a data domain level, decides on day-to-day topics, and ensures proper knowledge management and knowledge exchange.

  (4) The *Master Data Operational Community* communicates global process changes and clarifies operational issues related to MDM.

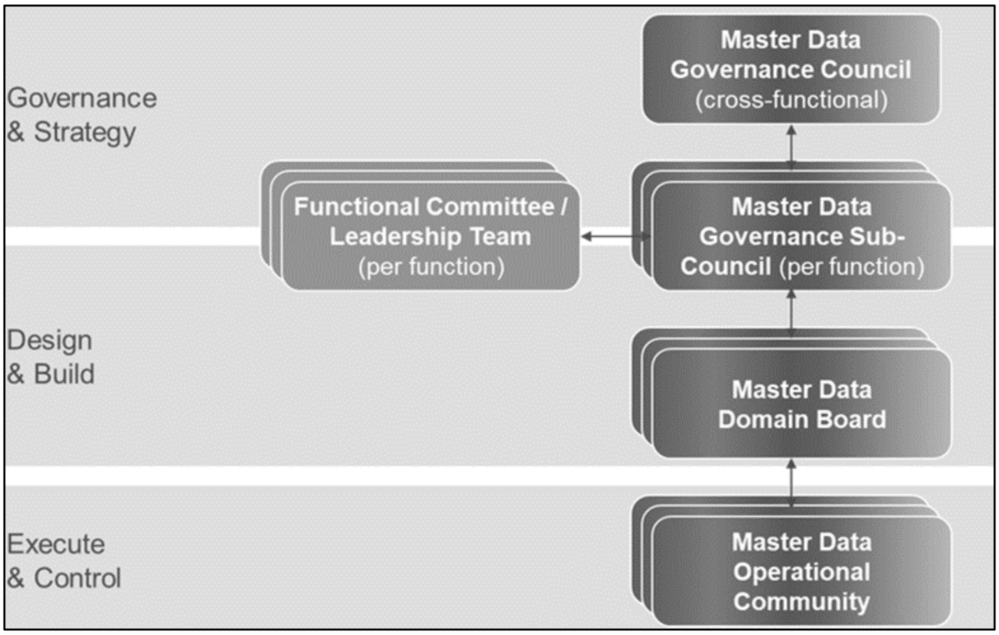


Figure 5-8: MDM boards at Bayer

- **Data Lifecycle** specifies the master data objects and documents for every relevant object its data sources, operational data activities, consumers, and use contexts. The data lifecycle describes the phases of data sourcing, creation, maintenance, transformation, release, archiving, and deletion. It is defined by the data owner, who aligns it with the consuming business processes, ensures its compliance with internal and external (regulatory) requirements, and takes care of the processes' continuous improvement.
- **Data Architecture** defines the conceptual data model, specifies which data is stored in which application, and defines how data flows between applications.
- **Data Applications** defines the software components supporting all MDM activities at Bayer. The Master Data Governance Sub-Council defines the functional requirements for these applications, while Bayer's IT department selects appropriate applications and takes care of their maintenance.
- **Performance Management** defines the measures to monitor and control the progress and outcome of data management with the help of a KPI system. Each Functional Data Steward develops and maintains a measurement system to monitor and control adherence to the directive's requirements.

Bayer's new corporate MDM directive became effective in April 2018. To communicate the contents of the directive, and to ensure transparency with regard to its links to other corporate directives, policies, and guidelines, Bayer has implemented a wiki-based solution for its MDM documentation. The starting page of the wiki depicts the DXM (see *Figure 5-9*).

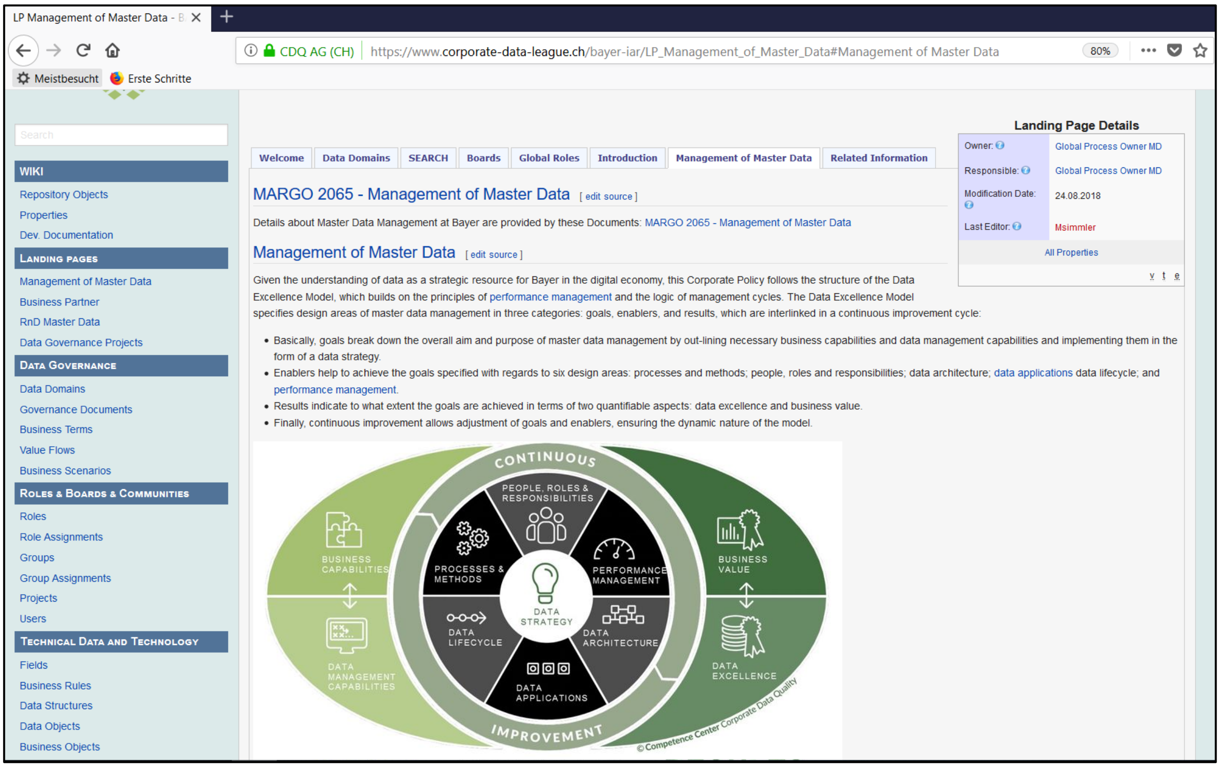


Figure 5-9: Screenshot of Bayer's MDM wiki

Lessons Learned

Developing a corporate MDM directive that is based on the DXM gives proof of the reference model's applicability and practical utility for defining and structing data management. The process design activities at Bayer indicated the need for enhancing existing MDM process reference models to incorporate all aspects of the DXM. The extensive discussions about the data lifecycle during the directive's development confirmed the previously made design decision to make "data lifecycle" a design area of its own.

5.4.3 SBB 1: Developing a Governance Policy for Infrastructure Data Management

This subsection documents the development of a corporate policy for MDM at SBB. The activities described in the following were conducted between April and June 2017 – prior to the development of the maturity model based on the DXM at SBB (see *Subsection 5.5.1*).

Company Profile

Swiss federal railway company SBB AG was founded in 1902. Its operations are split into four divisions: passenger traffic, real estate, freight services, and infrastructure. In 2018, SBB conveyed a daily average of over 1.2 million passengers and 205,000 tons of goods from and to 793 train stations (SBB, 2019) (see *Table 5-6* for details).

Table 5-6: SBB company overview

Schweizerische Bundesbahnen (SBB)	
Founded	1902
Headquarters	Bern, Switzerland
Industry	Transportation
Revenues	9.6 billion CHF (2018)
Earnings before interest and taxes (EBIT)	0.6 billion CHF (2018)
Number of employees	32,309 (2018)

Initial Situation

In mid-2017, SBB announced its “Data Strategy 2017–2020” aiming at managing data as a corporate resource. The strategy consists of three data management related objectives (i.e. ensure data quality, ensure common usage, and ensure continuous development) and four business value oriented objectives (i.e. build trust, create added value, enable innovation, and optimize business processes) (see *Figure 5-10*).

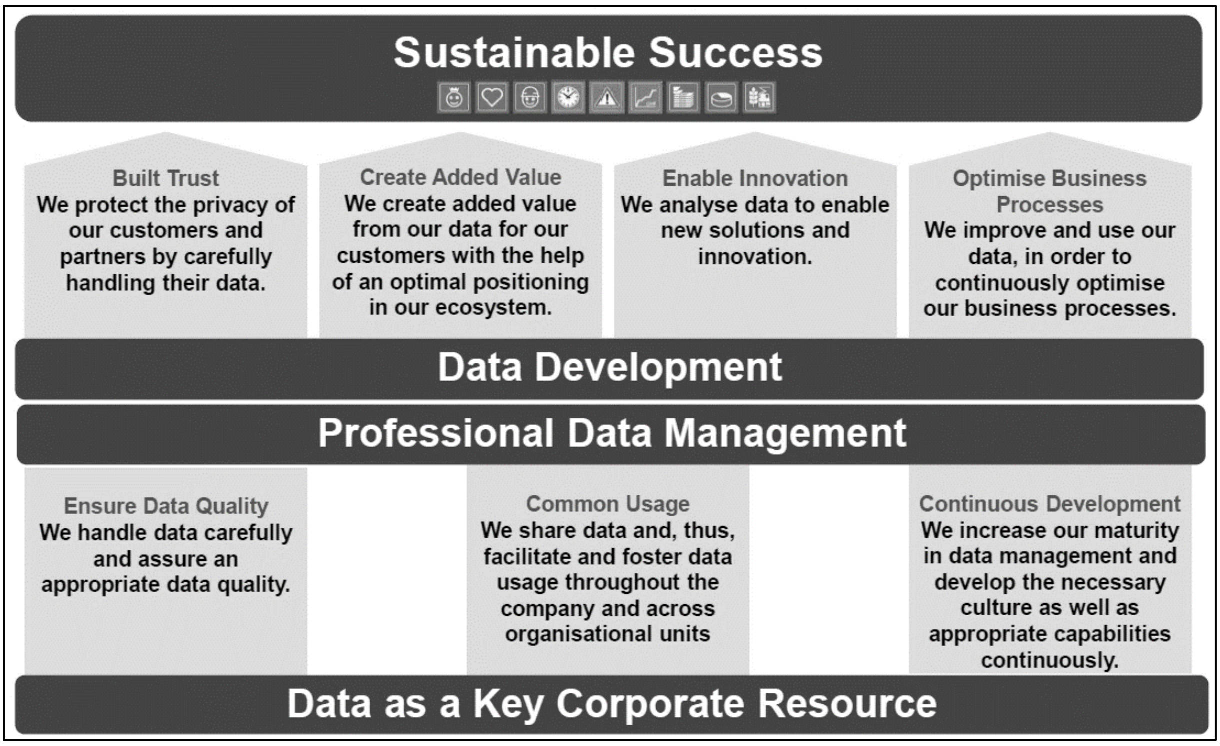


Figure 5-10: SBB's data strategy 2017-2020 (SBB, 2017a)

The data strategy explicitly refers to and depicts the DXM as the guiding framework for all data management activities at SBB. Eleven leading data managers are now responsible for operationalizing the data strategy in their respective data domain. For the data domains of the infrastructure division (i.e. fixed assets, topology), the leading data managers decided to update an earlier governance policy, which had been issued in a first version in 2013, but did not reflect the new corporate data strategy. This policy was intended to standardize the data management organization, provide guidance for employing the data management role model (which had only been partially implemented before), and define a target picture for the data management-related IT application landscape.

Solution

The author of the dissertation was asked to support SBB's data management in the infrastructure division in developing an appropriate governance document. A key decision was made at the beginning of the governance policy development process: The document should be structured according to the DXM, which is defined by SBB's data strategy as the enterprise-wide framework for data management. To do so, the author, together with a data manager from SBB, translated the English DXM terminology into a German, SBB-specific version of the reference model (see *Figure 5-11*). While the majority of design areas simply needed to be translated, SBB wanted to change the wording of four design areas in order to ensure compatibility of the reference model with SBB's terminology and corporate culture: "People, Roles, and Responsibilities" was changed into "Organization", "Data Applications" into "IT Products", "Data Architecture" into "Data Model", and "Performance Management" into "Measuring Systems".

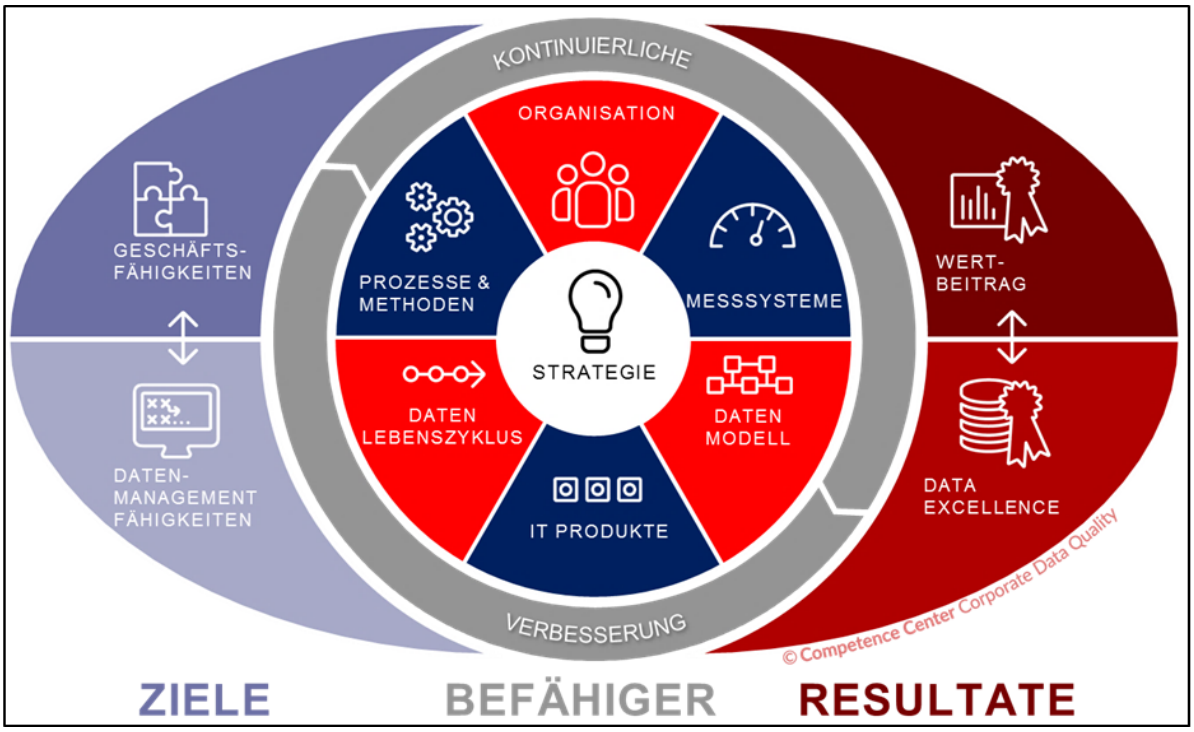
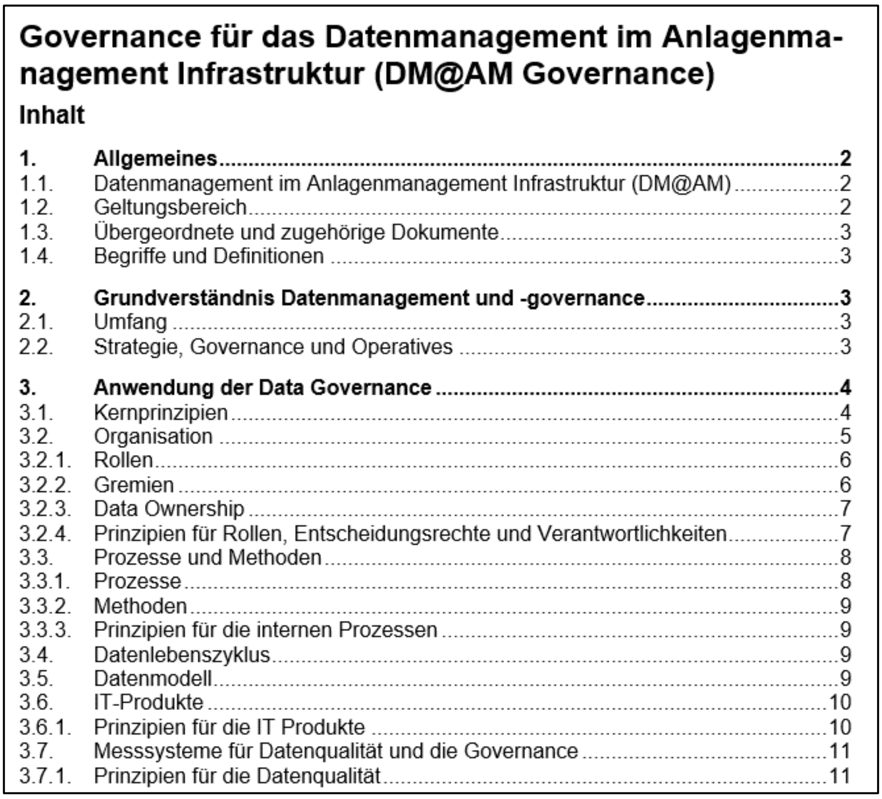


Figure 5-11: SBB-specific Data Excellence Model (in German) (SBB, 2017b)

As SBB's corporate data strategy explicitly addresses the design areas of the Goals and Results sections of the DXM, it was decided that the focus of the governance policy should be on the Enablers section of the model. Thus, after two introduction chapters, the core chapter of the policy outlines guiding principles for data management and specifies the six Enablers design areas (see *Figure 5-12*):

- **Organization** defines data management roles and boards based on an SBB-specific adaptation of Weber (2009) reference model for a data governance organization. Furthermore, the design area details the concept of data ownership and outlines organization-related principles (e.g. push-communication; each role has to actively communicate relevant or new information to its stakeholders).

- **Processes and methods** defines data management processes based on Reichert's (2015) reference model for MDM processes, and provides a list of, and links to, related policies, tools, and documents.
- **Data lifecycle** defines the relevant phases from data creation to archiving to deletion.
- **Data model** introduces and defines six modeling principles: (1) documentation of relevant data in a conceptual data model, (2) clear definition of data and documentation of metadata, (3) harmonization of enterprise-wide attributes, (4) central documentation of metadata, (5) enterprise-wide integration of data, and (6) integrated, end-to-end data flows.
- **IT products** defines requirements for managing the application landscape.
- **Measuring systems** outlines the relevant dimensions of data quality (i.e. completeness, timeliness, accuracy, and consistency) and specifies requirements for a technical data quality cockpit.



Governance für das Datenmanagement im Anlagenmanagement Infrastruktur (DM@AM Governance)

Inhalt

Figure 5-12: Table of contents of SBB's governance policy for infrastructure data management (in German)

Lessons Learned

The development of the governance policy at SBB gives proof of the applicability and practical utility of the DXM for structuring and designing data management activities and developing supporting documents. Furthermore, the case study illustrates how the reference model can be instantiated to fit a company-specific context (i.e. translation into German and adaptation to SBB's terminology). Developing the policy required to further detail the constructs of the reference model's design areas (e.g. referencing to Weber's (2009) reference model for a data governance organization or Reichert's (2015) reference model for MDM processes).

5.4.4 Schaeffler: Extending Data Management to Other Data Domains

Company Profile

Schaeffler is a global, integrated automotive and industrial supplier. The company produces high-precision components and systems for engine, transmission, and chassis applications, as well as rolling and plain bearing solutions for many industrial applications. With more than 90,000 employees worldwide, Schaeffler is one of the world's largest family owned companies. It has a worldwide network of manufacturing locations, research and development facilities, and sales companies in more than 150 locations in over 50 countries (Schaeffler, 2019) (see *Table 5-7* for details).

Table 5-7: Schaeffler company overview

Schaeffler	
Founded	1982
Headquarters	Germany
Industry	Automotive
Revenues	14.2 billion EUR (2018)
Earnings before interest and taxes (EBIT)	1.4 billion EUR (2018)
Number of employees	92,478 (2018)

Initial Situation

As a production company, Schaeffler has always been striving for high process efficiency and high-quality process outcome – and thus for high master data quality. Schaeffler started its MDM activities in 2009, with its central Corporate Data

Management (CDM) function being in the lead. An initial assessment and benchmarking project in 2014 allowed Schaeffler to assess the maturity of its MDM capabilities and identify areas with need for action. The results revealed MDM capabilities of high maturity in many respects, but also considerable potential for improvement. These areas were then systematically addressed, leading to significant improvements: an MDM strategy has been put in place and is being extensively communicated; performance indicators have been defined to measure and improve data and process quality; roles, responsibilities, and processes have been defined and implemented; data models and metadata models have been established; and applications are being managed centrally. All these activities have led to tangible benefits for Schaeffler's business processes, such as more clarity and commitment for business stakeholders (including a reduction of service processing time of 60 per cent), a comprehensive KPI methodology to continuously improve data quality (98,8 percent within the "debtor" data domain, for example) measuring CDM's business impact. So overall, MDM activities at Schaeffler had reached quite a high level of maturity, which resulted in winning the CDQ Good Practice Award for excellence in data management in 2016.

Having achieved this maturity with regard to master data, CDM at Schaeffler intended to broaden its scope and apply these proven methods for MDM to other data domains as well. A special focus was put on "Industry 4.0", one of Schaeffler's key concerns for the future. Among other things, Industry 4.0 aims at an efficient industrial production of customizable products (Hermann et al., 2016), in which data generated by machines and their sensors plays a key role. Monitoring the condition of machines is seen as one of the basic capabilities for implementing Industry 4.0 scenarios. CDM reached out to their colleagues from the Digital Coordination Team and from manufacturing (especially those responsible for machine condition monitoring in Schaeffler's facilities). Until then, the central data managers had not had any contact to these stakeholder groups. The people from the Digital Coordination Team and the machine experts agreed to join the CDM team in a workshop in order to get a better understanding as to whether and how data management methods could support them in their overall condition monitoring activities and, particularly, in managing sensor data.

Unplanned machine downtime has an immediate negative impact on production processes. Consequently, Schaeffler aims at preventing downtime in its production facilities by continuously monitoring the machines' conditions, deriving recommendations for improved operations and maintenance, and initiating and controlling implementation of these recommendations. In one of Schaeffler's production facilities, machine monitoring is conducted in five different ways: vibration analysis, thermography, fluid management (e.g. measuring oil moisture or counting oil particles), energy monitoring (e.g. electricity, voltage, active power), and operation analysis (e.g. stroke rate, pressure). The sensors, which are either installed inside the machines or mounted directly on the machines' housings, measure up to 250 values per machine. The machines are monitored in intervals ranging from every second to every 15 minutes. Software agents have multiple interfaces for different machine and sensor vendors and consolidate data from approximately ten machines. The monitoring activities for currently about 150 machines result in about 30 megabytes of newly generated data every day. Based on this data, visual maintenance dashboards are generated. In case of deviations, service and spare part orders are automatically initiated and email alerts are sent to service personnel. The machine monitoring activities in this plant are managed and supported by multiple functions and stakeholders on various levels. These include plant technology maintenance, industrial engineering, procurement and logistics, as well as one person that is responsible for all sensor data, and several users of this data at the facility level. Furthermore, corporate IT and operations technology support the initiative on a corporate level. External partners – such as machine vendors and their suppliers, subsystem suppliers, or manufacturers – are also involved in the condition-based machine monitoring activities. These activities started in 2006. Today they cover about twelve percent (approximately 150 of round about 1250) of the machines in the production facility. Due to the promising results, Schaeffler plans to successively increase the number of machines to be monitored. In addition, Schaeffler is designing a standardized condition monitoring system platform for a more scalable solution.

Solution

The author of this dissertation developed a questionnaire based on the DXM to discuss key aspects of data management with non-experts from the Digital Coordination Team and from manufacturing at Schaeffler (the questionnaire is depicted in *Appendix C.2*). Experts from Schaeffler's business units and functional areas, the Digital Coordination Team, and CDM participated in two workshops, which were moderated by the author of this dissertation together with a consultant from CDQ AG. The workshops focused on the condition-based machine monitoring activities in one of Schaeffler's production facilities (see above). At the beginning of the first workshop, the author of the dissertation explained to the participants the workshop's objectives and the rationale and content of the DXM. After that, the production facility experts laid out their activities for condition-based monitoring. In the remainder of the first and during the second workshop, the questionnaire was discussed and answered by Schaeffler's representatives leading to the following insights and recommendations for Schaeffler.

To enable condition-based machine monitoring, five **Business Capabilities** and eight supportive **Data Management Capabilities** are required (see *Table 5-8*). From a business perspective, the management of machine statuses (business capability 1), the triggering of business processes (3), and the presentation of condition-based insights (4) are on a sufficient maturity level. Nevertheless, there is room for further improvement. The next steps will focus on the development of a scalable, standardized condition monitoring system platform and the expansion of the knowledge base (2) to support the goal of "turning data into action" (5) as effectively as possible. From a data management perspective, a common terminology, a data model and metadata management are the basis for machine condition monitoring (see data management capabilities a, c, f, h), as multiple data types and data domains are of relevance.

Table 5-8: Business and data management capabilities for condition-based machine monitoring at Schaeffler

Business Capabilities	Data Management Capabilities	Enablers in the Data Excellence Model
1. Manage machine statuses (define statuses which may lead to machine downtime)	a. Define, maintain, harmonize and provide machine status value lists and the accompanying metadata (i.e. required roles, responsibilities, processes)	Data architecture; People, roles & responsibilities
2. Learn from experiences and share knowledge (make one mistake only once)	b. Structure and provide insights c. Define a common vocabulary	Data lifecycle; Data architecture; Processes & methods
3. Trigger business processes (e.g. QS alert, order spare parts)	d. Document events and activities e. Trigger processes (based on events)	Data lifecycle; Processes & methods
4. Report and visualize condition-based insights	f. Manage further business requests (what does the data user want?), define the appropriate solution (which "information services" are required?) and the required data (which data is needed from which sources?) g. Define and implement processes for gathering, enriching, combining and providing data (sensor data lifecycle management)	Data architecture; Data lifecycle; Processes & methods
5. Plan and conduct maintenance activities (turn data into action)	h. Define "maintenance events" as a data domain and manage its governance	Data architecture; Data lifecycle; People, roles & responsibilities; Processes & methods

People, Roles, and Responsibilities have only partially been defined and documented (e.g. no role descriptions for managing and using sensor data exist, no data ownership has been defined). Due to the small number of people currently involved, the business owners see a medium need for action in that area. However, the role and interaction model established by CDM is regarded as a sound starting base, which will be applied and adjusted to the data in scope.

Date management **Processes and Methods** are not being managed in a systematic way yet. Methods are predominantly developed "on demand" and not adequately documented. For example, responsibilities and rules defining when to apply which method are missing. Consequently, the business owners see a high priority to address this topic.

Data Lifecycle processes have partially been defined. Whereas data creation has been documented on a detailed level in the process landscape, processes referring to changes made to data, data maintenance, or data archiving or deletion have not been documented yet. Nevertheless, the need to detail the data lifecycle processes is regarded as relatively low, since most of the data is used only for a limited period of time after it has been created.

Relevant **Data Applications** for condition-based machine monitoring include a content management system (CMS) database, a CMS platform, and a plant maintenance application. The data from the various sensors and devices is collected by an "autinity agent", which has multiple interfaces and is stored in the CMS database. The CMS platform processes and analyzes the data, and reports and visualizes the insights. In case of alerts, the CMS platform triggers activities in the plant maintenance application. Until now, no direct connection to the manufacturing execution system (MES) has been established. Data can be exported either by a specific tool or via a CMS database interface. The business owners are satisfied with the current data application landscape.

Data Architecture includes various physical data models in the previously described applications. A common, application independent data model does not exist. Furthermore, metadata, such as data ownership and access rights, is currently not being documented and managed. In some cases, the measurement system for some data attributes (e.g. degree Celsius or Fahrenheit) has not been defined. The business owners see a need to establish a basic, application independent data model and metadata repository.

Some form of **Performance Management** for controlling data quality and data security does not exist – apart from a number of filters and alert rules in the CMS platform. The business owners see the absence of a performance management system as the most important challenge, because automatic alerts and initiations of maintenance activities rely on data quality. To define maintenance events and analyze past causes of machine downtime, data quality needs at an adequate level for an extended period of time. In addition to data quality, data security plays a prominent role, as machine data is of high relevance for efficient production processes and, thus, business critical.

Data Excellence in the context of condition-based machine monitoring has a strong focus on data quality. Especially correctness, completeness, consistency, and accessibility are important quality dimensions for machine monitoring at Schaeffler. Data security is another relevant aspect of data excellence.

The **Business Value** provided by condition-based machine monitoring and the supporting data management capabilities mainly refers to product quality, process efficiency, and supply reliability (through prevention of unplanned machine downtime). Moreover, early detection of machine defects prevents secondary damage, results in planned repair activities (at best, in parallel with regular maintenance activities), and reduces inventory cost for spare parts (which can now be ordered on demand).

The workshop results proved the general importance of, and need for, corporate-wide data management, as digital and data-driven activities are of cross-functional and cross-divisional nature and based on high-quality data. Management of machine data has been implicitly conducted – and without a coordinating structure or framework but with some design areas of the DXM partially addressed. The methods established by CDM were positively evaluated by the stakeholders of the condition-based machine monitoring initiative, and can be transferred to other (sensor) data types, which were previously not in scope of CDM. As next steps, data managers will be nominated for data domains classified as critical and relevant. The managers of these data domains will be responsible for developing a data domain strategy and transferring the well-established data management methods to their respective domain.

Lessons Learned

Applying the DXM in a "non-traditional" data management scenario confirmed the practical usefulness of the reference model – both as a means for structuring and documenting data-related activities and requirements and as a tool for discussing data management topics with business, IT, and other stakeholders in a systematic way. The design areas of the DXM can be considered valid regardless of the data types in scope (i.e. "traditional" master data or "new" sensor data). However, the design areas require some adaptation with regard to their content. This includes adding new roles, defining more flexible data ownership regulations for transactional data, designing data lifecycle processes for transactional data (e.g. maintenance is not required; instead, archiving is critical due to the large data volumes), and establishing architectural approaches to standardize (unstructured) data in order to process it. Finally, the case study illustrates that the questionnaire based on the DXM (see *Appendix C.2*) can serve as an instrument to extend data management activities to new data domains.

5.4.5 Bosch: Developing a Data Strategy

Company Profile

Robert Bosch GmbH was founded in 1886. Today, it is one of the leading automotive suppliers in the world. Its automotive division called "Mobility Solutions" generates more than half of the company's revenues (i.e. 47.4 of 78.1 billion euros in 2017). Further divisions are Industrial Technology, Consumer Goods, and Energy and Building Technology. Bosch has more than 440 subsidiaries in 60 countries throughout the world (Bosch, 2018, Bosch, 2019) (see *Table 5-9* for details).

Table 5-9: Bosch company overview

Bosch	
Founded	1886
Headquarters	Stuttgart, Germany
Industry	Automotive, industrial, consumer goods, energy, building technology
Revenues	78.1 billion EUR (2017)
Earnings before interest and taxes (EBIT)	4.9 billion EUR (2017)
Number of employees	402,166 (2017)

Initial Situation

For Bosch, the Internet of Things (IoT) and artificial intelligence (AI) are topics of high strategic relevance. The company aims at connecting all of its electronic products with each other by 2020 and applying AI for all electronic products in their development, manufacturing, or usage by 2025 (Bosch, 2019). In the implementation of Bosch's strategy of becoming the leading AI-driven IoT company, data plays a prominent role. However, while several data strategies had already been in place for specific business units or initiatives, a corporate-wide approach was still missing. Consequently, Bosch's Chief

Digital and Technology Officer in early 2019 decided to develop a data strategy, which should describe how data is to be gathered, stored, processed, and managed, and how data should generate value for Bosch's (internal) operations and (external) business activities.

Solution

To develop the data strategy, the Chief Digital and Technology Officer nominated two teams – one within the Central Enterprise Architecture team and one within the company's AI team – to draft a data strategy. This complementary approach resulted in different perspectives on data management and usage, which were combined into a single data strategy. This data strategy will be developed further in coordination of the Central Enterprise Architecture team.

The Central Enterprise Architecture team defined three layers (strategic framework, objectives and guiding principles of design areas, and implementation) as the guiding structure of Bosch's data strategy, using the six design areas of the DXM's Enablers section (see *Figure 5-13*). For developing the data strategy, the team followed a use case driven approach. To do so, the team identified and documented more than ten data-related use cases during workshops with relevant stakeholders from various divisions, business units, and initiatives. The use cases covered different views of data and different usage scenarios (e.g. management of person-related user data, manufacturing reporting across plants, self-service analytics, or training data lifecycle optimization in connection with machine learning). To analyze each use case, the team documented its general idea and data-related requirements, before it derived the capabilities required for the six enabling design areas of the DXM (e.g. use case "management of person-related user data" required a conceptual data model for consent management in the Data Architecture design area). By aggregating and prioritizing the capabilities required in all use cases, the team was able to specify the design areas, scope and value contribution of the data strategy. In parallel, the team reviewed Bosch's business strategy and IoT strategy and derived requirements to be met by data from a strategic perspective.

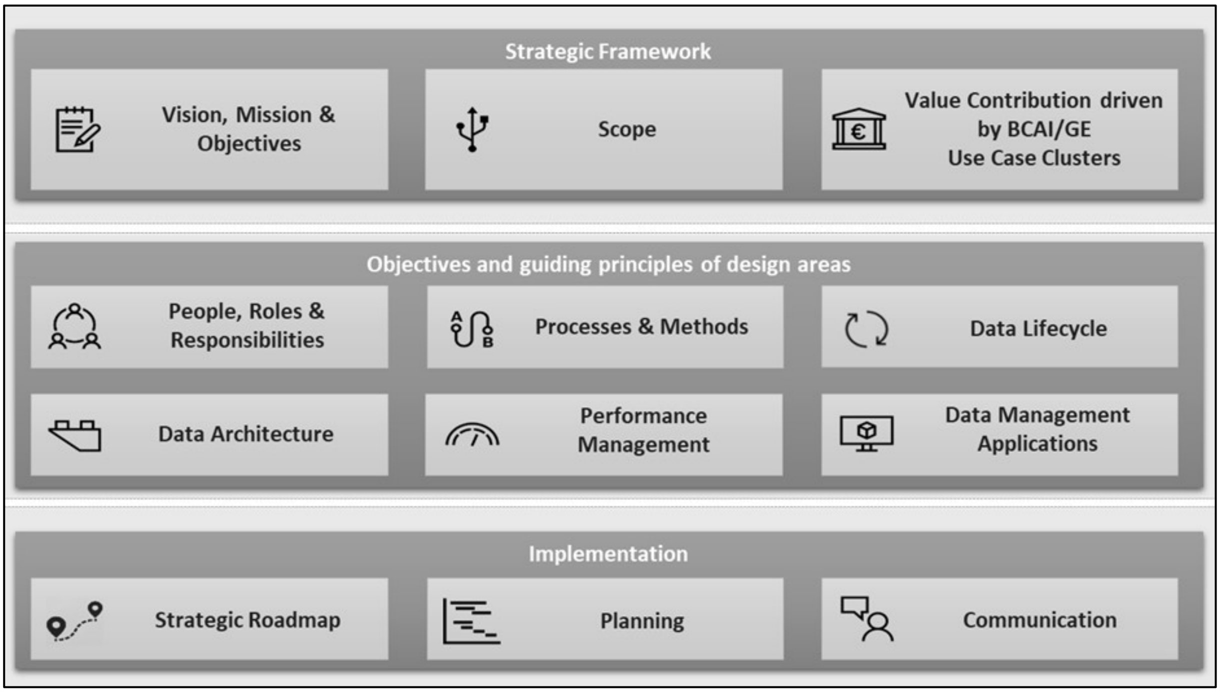


Figure 5-13: Bosch's Data Strategy Framework

Strategic Framework

Bosch's data strategy includes a vision and a mission statement, as well as four strategic objectives, which address either defensive or offensive elements of Bosch's data activities:

- Vision: *"Turn data into a key asset for RB*[35] *to become the leading AI-driven IoT company*"
- Mission: *"Data becomes a strategic asset, fit for use and marketed RB wide"*

[35] Internally, "(Robert) Bosch group" is referred to as "RB".

- *Defensive objectives*:
  - Compliance: *"We establish standards for data usage and provisioning to create transparency, assure legal compliance and minimize legal and operative risks."*
  - Trust: *"We establish data rules and policies that provide the best data quality, security, and control for our customers to establish RB as the most trustful IoT partner."*
- Offensive objectives:
  - Value: *"We establish and maximise* [sic] *the data value contribution for data-driven and data business processes, e.g. decision making, analytics, automation, enhance [sic!] customer services, …"*
  - Innovation: *"We enable new ways of utilizing our data for new business models, partnerships, and future technologies like artificial intelligence."*

The data strategy has a group-wide scope (i.e. Robert Bosch GmbH and all its subsidiaries) and addresses all data that is created, (externally) acquired, managed, or used. Regarding value contribution of the data activities, the strategy development team calculated a significant value potential of the use cases.

Objectives and guiding principles of design areas

For each of the six design areas, the team formulated first drafts of guiding principles, defined and specified core elements (i.e. the resulting deliverables), and listed the use cases which are enabled by the core elements of the design area. For People, Roles, and Responsibilities, for example, the team defined six guiding principles, among them *Foundation & Alignment* ("Basic roles are defined in alignment with international frameworks, e.g. CDQ, DAMA, as well as existing role descriptions at Bosch, e.g. MDM, AE, ISP, giving clear tasks and responsibilities.") or *Continuous Development* ("Central and Business wise Data Management Offices providing training and support of processes, best-practices and tools"). Furthermore, the data strategy lists three core elements related to this design area: (1) role model, (2) operating model, and (3) training concept:

(1) The role model defines the basic roles required to cover the main processes and responsibilities for the management and usage of data. Roles specified are Chief Data Officer, Data Owner, Data Possessor, Data Steward, Data Custodian, and Data Consumer.

(2) The operating model explains the patterns of interaction between these roles. It embeds the roles in the organization and provides an overview of boards required.

(3) The training concept ensures that the role owners are equipped with the right skills and competencies to fulfil their specific tasks.

Finally, the data strategy names the prioritized use cases enabled by the People, Roles, and Responsibilities design area and illustrates how these use cases are supported.

Implementation

The implementation layer of the data strategy details how Bosch intends to implement its strategic data vision. It outlines the strategic roadmap (which follows a use case driven approach, starting with the use cases generating the highest value), plans and details the required actions per activity block, and specifies the means of communication.

Lessons Learned

While data strategy development at Bosch was influenced by the DXM (especially by the six design areas of its Enablers section), the use case in turn confirmed several areas, aspects, and elements of the reference model as detailed in this dissertation. First and foremost, the strategy framework applied by Bosch confirmed the structure and elements of the Data Strategy Canvas (see chapter 4.3.2), which was developed around the same time but without Bosch's involvement. In addition, the fact that Bosch's data strategy adopted the six design areas of the Enablers section of the DXM confirmed the validity and relevance of these design areas. Furthermore, Bosch's specification of the six design areas (i.e. the formulation of guiding principles and core elements) provides evidences for success criteria, recommended practices, and key result documents.

5.5 Case Studies Applying the Generic Reference Model

This section outlines two case studies (conducted with SBB and tesa) illustrating how the reference model is applied as abstract design knowledge for communication, education, maturity assessment, and benchmarking purposes.

The CDQ Academy is another example of applying the DXM as a generic reference model. However, as this example does not address concrete actions within a company, it cannot be considered a case study comparable to the other ten cases presented in this dissertation. This is why the CDQ Academy example is documented in *Appendix A.4* for the purpose of additional illustration of the model's applicability.

5.5.1 SBB 2: Developing a Maturity Model

Initial Situation

By 2017, SBB had been using the Maturity Model for EDQM for selected master data domains for several years already. That year, SBB developed and announced its "Data Strategy 2017–2020", together with an activity roadmap. Among other things, the roadmap included the plan to assess the maturity of four master data domains (supplier, private customer, material, and topology) in 2018, and define target values for 2020. In addition, data managers from all other SBB data domains (i.e. business customer, rolling stock, schedule, employees, disruption, real estate, assets) were invited to also conduct maturity assessments. In previous years, a consulting company had supported SBB in conducting questionnaire-based maturity assessment interviews (typically about ten interviews per domain) and deriving a maturity score. With maturity assessments being defined an important element of the new data strategy, SBB intended to build up own capabilities for conducting maturity assessments in its corporate data management department.

Following the development of the DXM, SBB was the first member company of the CC CDQ to ask for a revised version of the maturity model. In an expert interview conducted in September 2017 (research activity 2-4 in *Table 4-3*), SBB data managers defined four key requirements to be met by such a revised version:

(1) *Structure and content:* Reflect the changing role of data and data management, and adapt the maturity model accordingly, by incorporating in the maturity model all relevant changes made to the initial Framework for CDQM in order to get to the DXM.

(2) *Continuity:* Ensure compatibility of the revised maturity model with the previous Maturity Model for EDQM to facilitate benchmarking with earlier assessment results.

(3) *Usability:* Ensure the questions for assessment are clear, unambiguous and easy to understand. To do so, incorporate feedback from users who conducted the assessment before. Furthermore, users of the maturity model should be supported by brief explanations and examples.

(4) *Simplicity:* Keep the maturity model as lean as possible (i.e. it should not contain more than the 38 questions of its predecessor, which were perceived as a sufficient).

Solution

Following the expert interview, the author of this dissertation designed a draft version of the revised maturity model, which was based on the Maturity Model for EDQM and the DXM. Intermediate versions of the model were presented to SBB's subject matter experts in two focus groups (taking place in October and November 2017) in order to collect feedback and optimize the model. In October 2017, the author used the maturity assessment questionnaire in two pilot interviews. In addition, two SBB data managers applied the model in interviews they conducted for assessing the maturity of the "topology" data domain. The author of this dissertation then conducted a final, SBB-internal focus group in December 2017 to collect feedback both from interviewees and interviewers, serving for further optimization of the questionnaire. In this focus group, the results of the maturity assessment, as well as the results from a benchmarking analysis against the assessment results from previous years[36] were presented to the participants, and experiences made with applying the model were reviewed and discussed. Finally, the author presented the revised maturity model in another focus group, taking place during a CC CDQ workshop in December 2017 (research activity 2-7). *Figure 5-14* gives an overview of these research activities.

In addition to the pilot assessment, SBB applied the revised maturity model in 2018 for the assessment of six data domains. For 2019, SBB has initiated and planned further assessment activities.

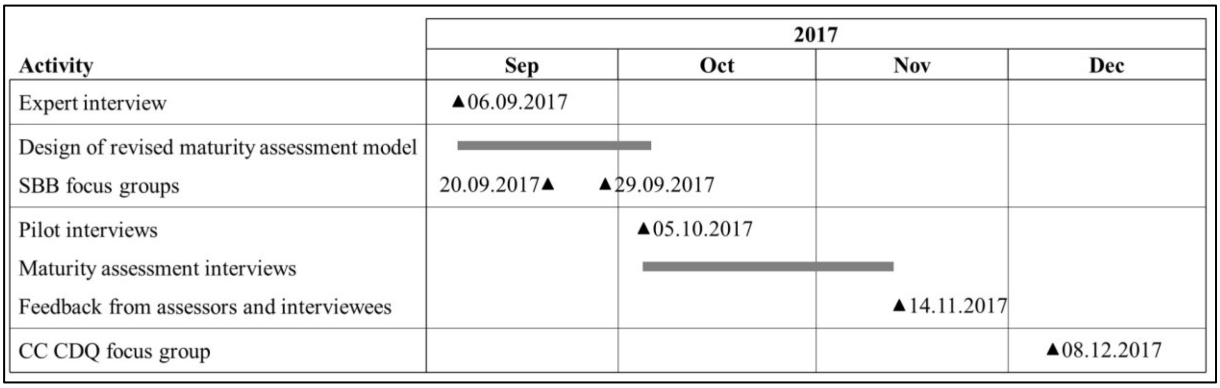


Figure 5-14: Design activities for revised maturity model

[36] The new maturity assessment model supports the conversion of previous assessment scores based on the Maturity Model for EDQM into the new scoring. This allows SBB to compare new assessment results with previous scores (cf. requirement 2 (continuity).

To address the requirement regarding continuity, it was decided to leave the assessment model of the Maturity Model for EDQM unchanged and use it also for the revised model. Consequently, the revised model is a continuous maturity model comprising five stages:

(1) *Not yet started:* No initiatives are in sight so far. Some valuable ideas exist; however, wishful thinking dominates.

(2) *Limited progress:* Some progress can be identified, like a successful implementation in a subdomain or a pilot project.

(3) *Average progress:* Basic approaches have been implemented in some areas. There are evidences of established procedures.

(4) *Significant progress:* Clear evidences for successful implementation exist. Initiatives have been implemented in almost all areas.

(5) *Fully completed:* Excellent and comprehensive results in all areas of the topic with high added value for users and stakeholders.

Furthermore, the scores of previous assessments can be entered into the new calculation template, allowing users to conduct a trend analysis over time. The other three design requirements listed above are addressed by the domain model of the revised maturity model. The maturity assessment questionnaire can be found in *Appendix C.1* of this dissertation.

- Structure and content: The domain model has the same structure and content as the DXM, with one section for the Goals of data management, six sections for assessing the Enablers of data management, one section for the Results, and one section concerning Continuous Improvement. Each section contains between two and six aspects to be assessed.

- Usability: The three focus groups conducted during the model design process included actual users of the maturity model. Furthermore, usability is ensured through a list of items indicating evidence for each statement.

- Simplicity: The domain model includes 36 statements for assessing the relevant aspects of data management.

SBB has applied the revised maturity model for eight data domains already (with more maturity assessments planned for the future). In addition, multiple other companies (among them ABB, Bayer, Beiersdorf, Schaeffler, Shell, Swarovski, and Zespri) have applied the maturity model based on the DXM (research activity 2-8).

Lessons Learned

The maturity model design activities allowed for a detailed review of the DXM since the DXM forms the basis of the maturity model's domain model. As continuity was a major requirement with regard to designing a revised maturity model, it was important for the author of the dissertation to create as much transparency as possible regarding the necessary changes that had to be made in order to get from the Framework for CDQM to the DXM. Consequently, the author highlighted the significant amendments to the Framework for CDQM, which had become evident in the form of additional statements in the domain model. This ensured the "downward compatibility" of the maturity and reference model with their predecessors.

Furthermore, the development of the domain model also enhanced the understanding of the DXM. The explicit request from SBB data management experts for a list of items explaining and illustrating each statement of the domain model resulted in a list of good data management practices based on a literature and case study review. Based on this list of good practices, the author of the dissertation identified typical deliverables of data management and mapped them against the meta-model of the DXM. This resulted in adding "data flow" as a deliverable of data lifecycle processes to the meta-model.

5.5.2 tesa: Communicating Data Management Activities

Company Profile

tesa SE is one of the world's leading producers of self-adhesive solutions for industry and trade, consumers, and craftspeople. The company has more than 125 years of experience in coating technology and the development of adhesive masses and innovative product solutions. tesa is a member of the Beiersdorf Group and became an independent corporation in 2001. Headquartered in Germany, the company has nearly 5,000 employees, six regional headquarters, eight plants, 65 offices, and 52 affiliates worldwide (tesa, 2019a) (see *Table 5-10* for details).

Table 5-10: tesa company overview

tesa	
Founded	2001
Headquarters	Norderstedt, Germany
Industry	Self-adhesive solutions
Revenues	1.3 billion EUR (2018)
Earnings before interest and taxes (EBIT)	0.2 billion EUR (2018)
Number of employees	4,917 (2018)

Initial Situation

In 2017, tesa established a central Corporate Data Management (CDM) team, aiming at providing a solid data foundation for the enterprise's business processes and operations, digitalization activities, and analytics and BI. CDM's mission comprises five major elements (tesa, 2019b):

(1) build up a data management organization,

(2) improve data creation and data maintenance activities,

(3) ensure high data quality,

(4) define and document tesa's data assets, and

(5) support projects and operations.

To establish a data management organization, the team used (Reichert, 2015) reference model for MDM processes for specifying a number of core data management activities. The team then defined roles and responsibilities (a) for the central CDM organization and (b) for decentral data management in the business departments. While central data management responsibilities were successfully taken over by members of the central organization, CDM faced the challenge of rolling out decentral data management responsibilities.

Solution

To address the need for contributions from business, the CDM team used multiple ways for communicating and promoting its goals and activities to the respective stakeholders. Beside extensive informal communication, the team made use of existing team meeting structures (such as department meetings or management meetings) to effectively position data management as a relevant topic. Furthermore, CDM established a data management community by organizing a bi-weekly "master data table" during lunch break in order to give all interested employees an opportunity to informally exchange opinions and views.

To be as effective as possible, CDM took advantage of a media mix allowing the team to inform all tesa employees about data management. Team members, CDM's mission, and its service offering were presented in the tesa company magazine (tesa, 2018, pp. 18–20) (see *Figure 5-15*). In this publication, the DXM, adapted to tesa's corporate identity in terms of color, played a prominent role: It is mentioned in the article, its graphical representation is depicted, and the six Enabler design areas are briefly explained in an info box.

18 TEAM CORPORATE DATA MANAGEMENT inside:tesa 3/18

Doreen Spange
Corporate Data Steward (Material)

She knows the operational handling of data in tesa's business like the back of her hand and is responsible for defining rules and guidelines for material data management.

UP TO DATE

Generating data and using them to successfully work across different functions and companies is much more than just a question of technology. The Corporate Data Management (CDM) team has begun creating a world of data based on a clear "language" that brings everything together.

Have you ever wondered what a product's color is considered to be? For a tape like tesafix® 4965, is it the tape itself (transparent) that counts, or the liner (salmon)? And what about a dispenser? Is the color defined according to the purple housing, or the transparent tesafilm® inside? That's all correct, and yet, from a data specialist's standpoint, it is imprecise: "To avoid delivering the wrong thing, we need to agree on what exactly is defined as the color in which field of each IT system. Our colleagues need standardized specifications for that," explains Jan Hinsch, head of the Corporate Data Management team, which was founded in April 2017. The task behind it might sound simple, but it is even more complex: to produce "clean," unambiguous master data at tesa.

HIGH DATA QUALITY ACROSS THE BOARD

In the summer of 2016, it became obvious that tesa needed to embark on corporate data management. The primary goal was to reach the same high standard of quality for data, across all systems and with an eye to being ready for the future. One of the initiators was Matthias von Schwerdtner, who was heading the Specific Industrial Markets BU at the time: "The more we exchange data internally, and most especially externally as well, with business and process partners like Internet platforms, logistics providers and customs authorities, the stricter the requirements when it comes to good data management. This is like the backbone of increasingly fast, interconnected, and globalized processes." With about 7,000 products and 40,000 item numbers in the tesa product range, the interfaces with customers and suppliers are another important factor: "Data management doesn't just stay within tesa; we also have to provide data formats with the right semantics for external systems," Hinsch explains, pointing to digital sales platforms such as Amazon, 1WorldSync, and Ariba Network, which are already important to tesa today.

Throughout the company, there are data produced both centrally and on a decentralized basis on various software platforms, from R&D to HR and from Supply Chain and Controlling to Customer Management. The challenge of developing a shared understanding and a shared language across all functions is not mainly a technical one, though. Instead, the crucial factors are who is responsible for which data and who has to be notified of changes. Hinsch offers a practical example: "Let's say tesa is introducing a new color. That color has to be stored uniformly in both the ERP system (SAP) and the product information system (PIM). So we need employees who keep things like this in mind and provide new data at the relevant points."

Jan Hinsch
Head of Corporate Data Management

He is responsible for the development and organization of corporate data management.

THE 6 ENABLERS OF CORPORATE DATA MANAGEMENT

People, roles & responsibilities defines the skills and organization to ensure effective data management and consistent use of data across the entire organization.

Performance management defines the measures to monitor and control the performance (i.e., progress and outcome) of data management with the help of a key performance indicator system.

Processes & methods defines procedures and standards for managing and using data properly and consistently.

Data architecture defines the conceptual data model, specifies which data is stored in which application, and describes how data flows between applications.

Data lifecycle defines data objects and documents, and reviews data sources, operational data activities (i.e., ranging from data acquisition and creation to data archiving), data consumers, and data use contexts.

Data applications defines the software components supporting data management activities.

inside:tesa 3/18 CORPORATE DATA MANAGEMENT TEAM 19

DIALOGUE WITH GLOBAL PLAYERS

This – the human factor – is exactly where the data management structure and practices to be developed at tesa start out. The basis is a data excellence model adapted to the company. The model consists of six strategic fields, starting with "People, Roles & Responsibilities" (see box). The model was developed by the Competence Center Corporate Data Quality (CC CDQ), a network that originated at the University of St. Gallen and consists of about 40 well-known companies in Central Europe and other representatives from

Heiko Langhein
Corporate Data Steward (Supplier)

A member of the tesa Purchasing Network, he is an expert on all supplier data.

Antje Schmidt
Data Architect

She is the "interface" with IT. She makes sure the data landscape harmonizes with tesa's IT systems.

Figure 5-15: Article in tesa's company magazine (tesa, 2018, pp. 18–20)

Furthermore, CDM produced an animated video illustrating the importance of data management in an entertaining and cheerful way (see *Figure 5-16* for a screenshot). Among other things, the video shows how incorrect data entries – visualized in the video as "gremlins" – have negative consequences for business processes (such as delayed delivery due to missing customs data, or delivery to a wrong location due to defective customer address data). The graphical representation of the DXM can be seen in several scenes showing the CDM team "cleansing" defective data.

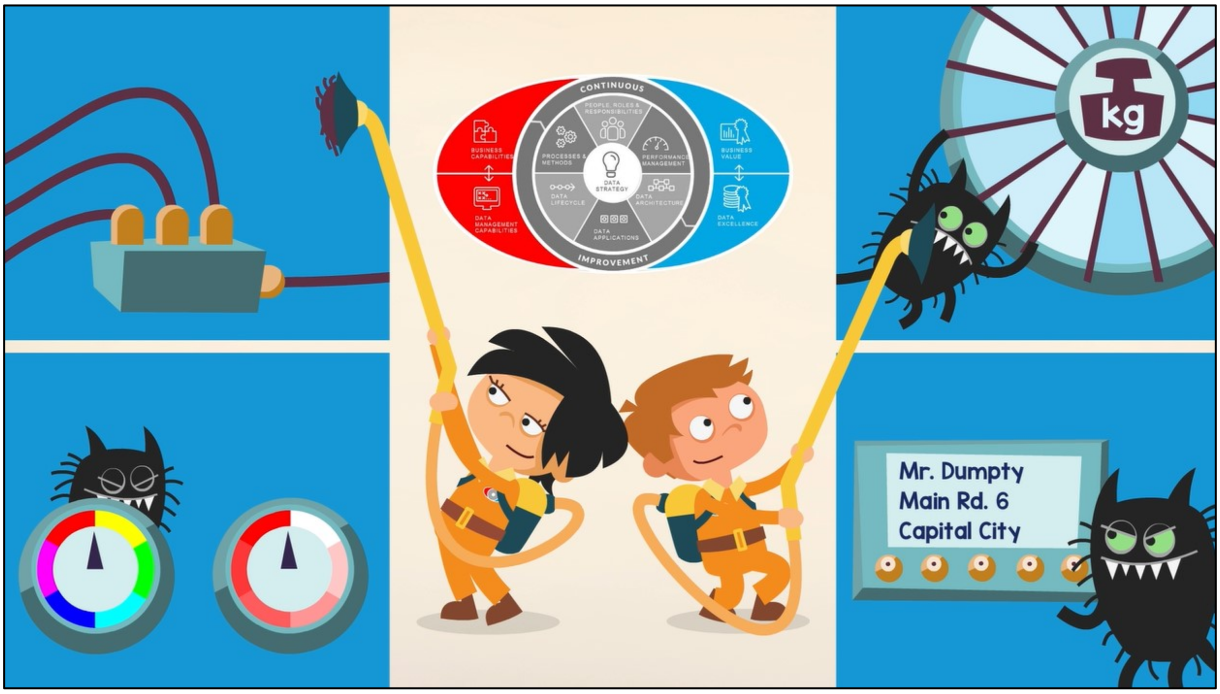


Figure 5-16: Screenshot of the animated video about tesa's corporate data management

Lessons Learned

The case study at tesa illustrates how the DXM can successfully be applied for various communication activities directed towards stakeholders from business. This gives proof of the reference model's applicability and practical utility for communication of data management activities to a broad audience of people who are not data management experts.

5.6 Cross-Case Summary

The ten cases presented illustrate both the problem space and the solution space of strategic data management. While the three exploratory case studies indicate the changing role of data and motivate the need for a capability reference model for strategic data management (i.e. problem space), the other seven case studies describe how the DXM, its instantiations, and applications can help tackle data management challenges (i.e. solution space).

The case studies presented provide sufficient insights regarding all twelve design areas of the DXM[37] (see *Table 5-11*). The number of case studies as well as the extent to which they cover the design areas indicate theoretical saturation, since each design area is supported by evidence from at least three case studies (Eisenhardt & Graebner, 2007; Glaser & Strauss, 2017).

Table 5-11: DXM design areas as covered by the case studies conducted

	Business Capabilities	**Data Management Capabilities**	**Data Strategy**	**People, Roles, and Responsibilities**	**Processes and Methods**	**Data Lifecycle**	**Data Applications**	**Data Architecture**	**Performance Management**	**Data Excellence**	**Business Value**	**Continuous Improvement**
1. PMI	X	X	X	-	-	X	X	X	X	X	X	X
2. SAP 1	X	-	-	X	-	-	-	-	-	-	-	-
3. SAP 2	-	-	-	-	X	-	-	-	X	X	-	-
4. Bayer 1	X	X	X	-	-	-	-	-	-	-	-	-
5. Bayer 2	-	-	X	X	X	X	X	X	-	-	-	-
6. SBB 1	-	-	X	X	X	X	-	X	-	-	-	-
7. Schaeffler	-	X	-	X	-	-	-	-	X	X	X	X
8. Bosch	X	X	X	X	X	X	X	-	-	-	X	-
9. SBB 2	-	-	-	-	-	-	-	-	X	X	-	X
10. tesa	-	-	-	X	-	-	-	-	-	-	-	-

Each case study contributed to the design of the artifact: The three exploratory case studies helped design the general structure (mainly the need for strong business orientation) and the (strategic) scope of the reference model; the other seven case studies helped define important details, provide feedback regarding the graphical representation of the

[37] The DXM and its twelve design areas are detailed in *Sections 6.2 and 0.*

artifact, and identify usage scenarios of the DXM. In line with the ADR methodology, the author's interaction with practitioners initiated further interaction with fellow researchers and other practitioners, leading to the final version of the artifact as described in the following *Chapter 6*.

6. Reference Model Design

This chapter outlines the design of the main artifact of the dissertation: the Data Excellence Model (DXM) as a capability reference model for strategic data management. *Section 6.1* provides an overview of the design requirements for the artifact. *Section 6.2* introduces the reference model by detailing its nature, structure, and meta-model. *Section 6.3* elaborates on the design decisions the author made. *Section 6.4* describes the data management roles specifically addressed by the DXM, and how each of these roles can benefit from using the DXM. Finally, *Section 6.5* describes each design area of the DXM on a detailed level.

6.1 Design Requirements

Building on the RBV and the understanding of data as an economic good, the DXM seeks to help organizations in managing data as a strategic resource. It mainly addresses global corporations, which often have distributed operations in complex organizational structures resulting in siloed data activities. In these corporations, establishing enterprise-wide data management is challenging as the complexity of data management increases the more an organization and its application landscape are distributed (Jain, Ramamurthy, Ryu, & Yasai-Ardekani, 1998).

Establishing data management significantly changes existing policies and practices and impacts headquarters, business lines, and every subsidiary (Haug & Stentoft Arlbjørn, 2011). From exploring and analyzing these challenges and reflecting the research questions and activities of the CC CDQ, the author specified six requirements to be met by a reference model for strategic data management (see *Table 6-1* for an overview). This specification was based on a review of past CC CDQ research activities (see *Section 4.2*), and by means of case studies (see *Chapters 5*), plenary discussions, focus groups, and expert interviews (see research activities in *Table 4-3*) with all findings being triangulated by reviewing academic literature.

- R1: Business orientation – In data-driven enterprises, data is becoming increasingly business critical. Besides technological capabilities, identifying and addressing a company’s data requirements demands close alignment of data management with business being the consumer of data. Thus, the reference model should help identify business-critical data requirements and derive responding data management activities.

- R2: Key constituents – The reference model should outline the key constituents of enterprise-wide data management. This means it should specify strategic, organizational, and technology aspects that are relevant for managing data.
- R3: Scope – Due to the emergence of smart factories, smart products, and social media, the number of data sources and the volume of data available is constantly growing. To make use of big data, generate data-driven insights, and data-based products and services, strategic data management must extend its traditional scope (which is to focus on master data) and include further data types, including analytical data, web data, and sensor data. Hence, the reference model should allow managing data that originates from multiple sources and is used for multiple purposes.
- R4: Purpose beyond data quality – In data-driven enterprises, data quality improvement is still considered a key goal of data management. However, further aspects gain relevance such as mitigating data-related risks and complying with increasing regulation relating, for example, to data privacy or data security. Thus, the reference model should contribute to the improvement of data quality and help address relevant data-related concerns.
- R5: Value contribution – In light of the digital and data-driven economy, the importance of data management has grown, and its scope keeps getting broader towards a business-critical capability. Therefore, the value contribution to business generated by data and data management should be made transparent by the reference model.
- R6: Implementation – In their data management activities, the CC CDQ member companies realized that building up data management capabilities was a very tedious endeavor. In fact, it took them several years to address and implement data management at an enterprise-wide level. To develop data management in a staged approach, understanding and assessing the current maturity of data management activities is beneficial. Thus, the reference model should incorporate the idea of continuous improvement and allow to assess the maturity of data management.

Artifacts developed according to DSR principles should be characterized by practical relevance and scientific rigor (Hevner et al., 2004). Regarding the DXM, *practical relevance* is ensured by addressing the requirements being directly derived from practitioners' needs. *Scientific rigor* of the DXM is ensured by the research design (see *Chapter*

4), which followed proven scientific standards for artifact design, good modeling practices, and general recommendations for IS artifact evaluation.

Table 6-1: Design requirements

No.	Requirement	Description
R1	Identify business-critical data requirements (*business orientation*).	Consider data requirements of digital business models and data-driven scenarios.
R2	Outline key constituents of data management (*key constituents*).	Specify the strategic, organizational and technology aspects relevant for data management.
R3	Manage data originating from multiple sources and being used for multiple purposes (*scope)*.	Include further data sources and data types in addition to master data.
R4	Address relevant data-related concerns *(purpose beyond data quality)*.	Consider data quality a key goal and key result of data management and address the relevance of compliance, data privacy, and data security as additional concerns.
R5	Demonstrate the business value generated by data and the value contribution of data management to business (*value contribution)*.	Create transparency regarding the business value of data and the value contribution of data management to business.
R6	Develop data management in stages (*implementation*).	Consider data management a long-term endeavor systematically developing over time. Understand the current maturity of data management activities to define improvement means.

6.2 Nature, Structure, and Meta-Model of the Reference Model

6.2.1 Nature

The DXM is a capability reference model for strategic data management. Taking a capability view, data management can be defined as the capability to successfully deploy data resources. The DXM builds on the understanding of data management as a *dynamic capability* that is contingent on business objectives (cf. Otto, 2012a, p. 14) and develops, combines, or reconfigures (data) resources as well as *core* and *ordinary data management capabilities*. This understanding materializes in the structure of the DXM:

- Capabilities are goal-oriented (Amit & Schoemaker, 1993). Accordingly, the DXM aims at explicitly documenting data's business critical role in the form of business capabilities, and then deriving core data management capabilities from these business capabilities, allowing data managers to directly align their activities to changing business requirements.

- Core capabilities are enabled by a combination of resources and ordinary capabilities (Stoel & Muhanna, 2009). The Enablers part of the DXM specifies a number of socio-technological design areas for developing the necessary core data management capabilities.
- Capabilities are results-oriented (Bharadwaj et al., 2013). Core data management capabilities seek to maximize business value by providing excellent data to the business.
- Capabilities evolve over time. Accordingly, data management is not a one-off effort, but an ongoing activity that is characterized by continuous improvement.

6.2.2 Structure

Figure 6-1 depicts the DXM as the main artifact resulting from the research process described in this dissertation. Based on six design decisions (see *Subsection 6.3*), the capability reference model comprises twelve design areas, which represent the main constituents (domains) of strategic data management:

- The design areas constituting the Goals section are (1) *Business Capabilities*, (2) *Data Management Capabilities*, and (3) *Data Strategy*.
- The Enablers section includes (4) *People, Roles, and Responsibilities*, (5) *Processes and Methods*, (6) *Data Lifecycle*, (7) *Data Applications*, (8) *Data Architecture*, and (9) *Performance Management* as design areas.
- The Results section consists of two design areas: (10) *Data Excellence* and (11) *Business Value*.
- All previous design areas are interlinked by (12) *Continuous Improvement*.

At its core, the DXM defines 30 success criteria and 93 recommended practices describing how excellent organizations conduct strategic data management. Each design area and its success criteria are briefly described in *Table 6-2* (for a detailed description of each design area, see *Section 6.5*). To support both capability-building and continuous improvement, the DXM specifies ambitions, success criteria, recommended practices, and key result documents to ensure each design area actually materializes. These findings are supported by justificatory knowledge from both the scientific and the practitioners' domain.

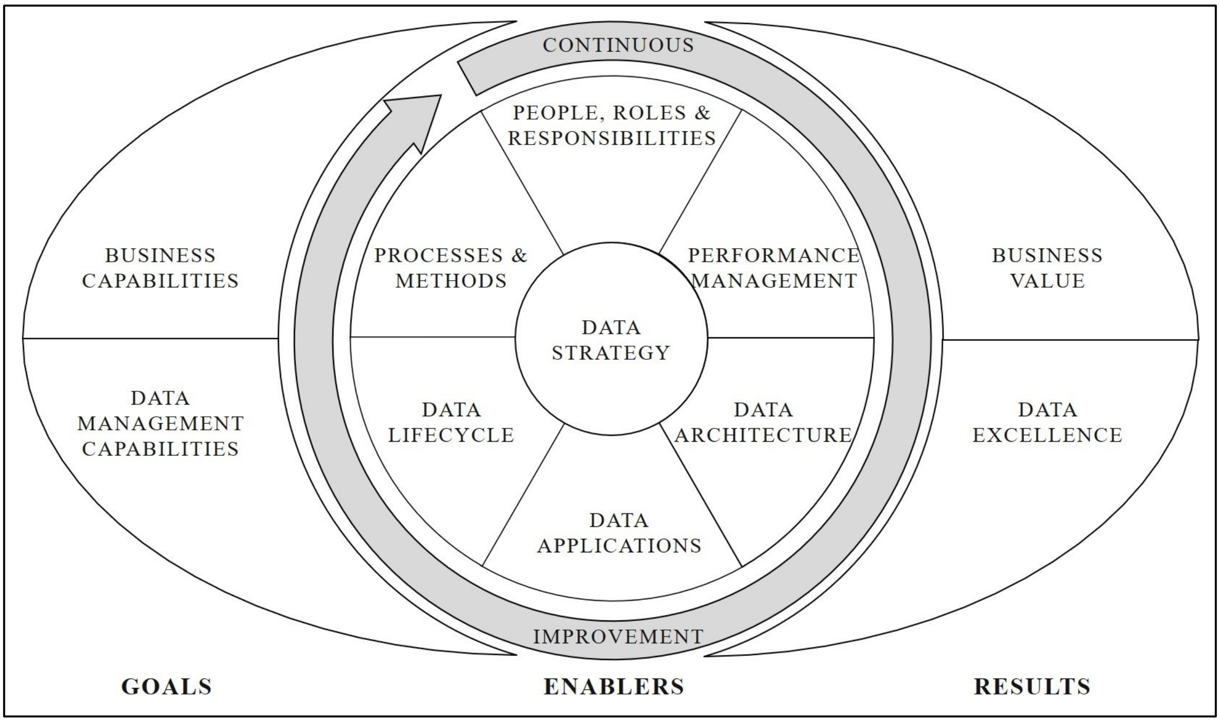


Figure 6-1: Data Excellence Model

Table 6-2: Design areas and success criteria of the Data Excellence Model

Design area (DA)	Description	Success criteria
DA1: Business Capabilities	Excellent organizations recognize the importance of data as a prerequisite to respond to business drivers and achieve business objectives. They exploit data for operational excellence, explore data-related business opportunities, and promote a data-driven working mode.	1a. Business requirements towards data are continuously analyzed and documented. 1b. New ways of using data in business models, business processes, and decision making are explored. 1c. The organization has a data-driven working mode.
DA2: Data Management Capabilities	Excellent organizations manage data as a strategic resource, provide data management services and excellent data to support data exploitation and exploration.	2a. Data provision and data management services are defined to support business capabilities. 2b. A portfolio of data provision and data management services is developed and maintained.
DA3: Data Strategy	Excellent organizations define and implement a data strategy, which specifies how to build a data foundation and how to use data in order to generate business value.	3a. Vision and mission for managing and using data are developed and communicated. 3b. Top managers are personally involved in defining and implementing the data strategy. 3c. The data strategy is integrated in the corporate management system.

Design area (DA)	Description	Success criteria
DA4: People, Roles, and Responsibilities	Excellent organizations maintain a data-driven culture, they develop, manage, and leverage their employees' potential regarding data on an individual, team, and organizational level.	4a. Roles, responsibilities, and decision-making rights in connection with data are embedded in existing structures, defined, and actively managed. 4b. A central data organization coordinates and facilitates data-related activities across the organization. 4c. Awareness of the importance of data is raised and continuously maintained across the entire organization. 4d. People are empowered to assume responsibility for data.
DA5: Processes and Methods	Excellent organizations develop and maintain processes and methods for efficient data management that fully satisfy diverse stakeholders' expectations and requirements.	5a. Data management processes and methods are designed, managed, and continuously improved. 5b. Data management activities are embedded in business processes, corporate policies, and standards.
DA6: Data Lifecycle	Excellent organizations define and efficiently manage all relevant data objects, their respective lifecycles, and use.	6a. Data objects and their lifecycles are unambiguously defined, systematically documented, and actively managed. 6b. Data use is traced, documented, and monitored.
DA7: Data Applications	Excellent organizations provide and maintain the applications required to efficiently manage and use data.	7a. Application landscape is planned, actively managed, and continuously optimized. 7b. Applications provide an integrated and consistent version of core data objects as an enterprise-wide reference ("golden record"). 7c. Applications provide functions to analyze, monitor, and continuously improve data quality. 7d. Applications provide functions to describe, catalog, provide, and access data and their metadata for consistent usage across the organization.
DA8: Data Architecture	Excellent organizations define core data objects through data models, document business rules, and design the data storage and distribution architecture for data being consistently understood and used across the entire organization.	8a. Common understanding of a data model for the business entities is developed, permanently assessed, and communicated. 8b. Data storage, distribution, and flow are designed, implemented, and actively managed. 8c. Business rules are systematically documented and managed.

Design area (DA)	Description	Success criteria
DA9: Performance Management	Excellent organizations define measures to continuously review and improve data excellence, performance, progress, and business value of all data management activities.	9a. Business value metrics are defined, actively managed, and measured. 9b. Data excellence metrics are defined, actively managed, and measured. 9c. Data management performance and progress metrics are defined, actively managed, and measured.
DA10: Data Excellence	Excellent organizations comprehensively report on data excellence achievements as well as on performance and progress of all data management activities.	10a. Data excellence is continuously reported and appropriately communicated to all stakeholders. 10b. Performance and progress of data management is continuously reported and appropriately communicated to all stakeholders.
DA11: Business Value	Excellent organizations comprehensively report on and achieve outstanding business value generated by data.	11. Business value generated by data and data management is continuously reported and appropriately communicated to all stakeholders.
DA12: Continuous Improvement	Excellent organizations regularly review the results of data management and adjust their enablers to further increase the level of data excellence and the business value generated.	12. Continuous improvement cycle encompassing all data management enablers is in place.

6.2.3 Meta-Model

A design area is ontologically defined by the entities (constructs) it addresses and the result documents it produces (representing outcomes of design activities). The design areas and constructs of the DXM as well as the interrelations between the constructs are specified by a meta-model (see *Figure 6-2*). Being a conceptual data model for strategic data management, the meta-model builds the ontological foundation of the DXM and creates a common understanding among experts from academia and practice (Hevner et al., 2004; Schütte, 1998). In the process of modeling the meta-model, the author of the dissertation followed the recommendations of the Object Management Group (OMG) and utilized the simplified Unified Modeling Language (UML) notation, providing aggregation and specialization relationships (OMG, 2017b). The meta-model is detailed by a glossary, which documents each construct, its definition, resulting documents, and justificatory knowledge (see p. 243 et seqq.).

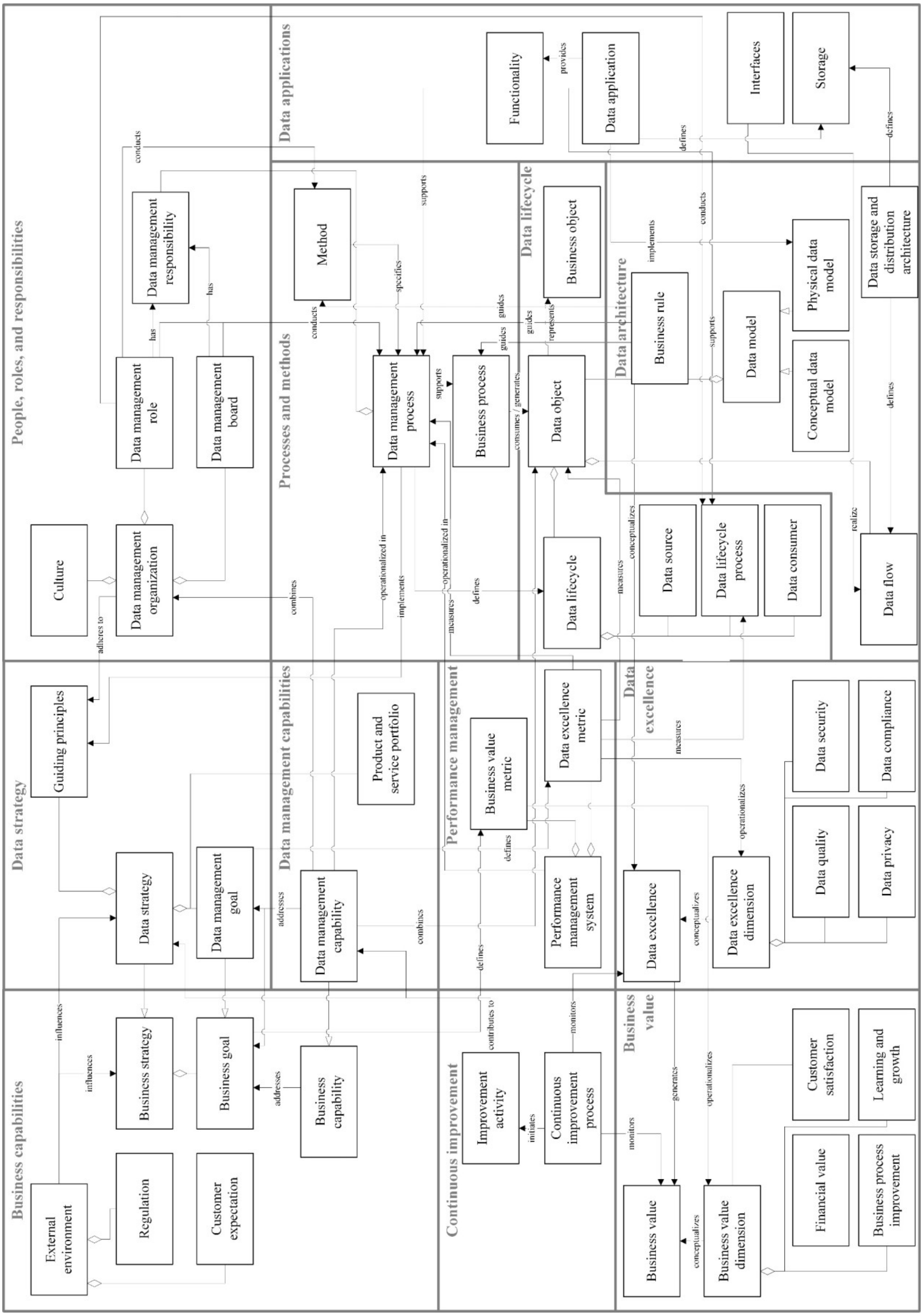


Figure 6-2: Meta-model of the Data Excellence Model

6.3 Design Decisions

As a consortium research artifact, the DXM incorporates the knowledge and experiences accumulated in previous phases of the CC CDQ (i.e. ontology and capability building for quality-oriented data management, see *Subsections 4.2.1* and *4.2.2*), while addressing the requirements of strategic data management at the same time (see *Subsection 4.2.3* and *Section 6.1*). Consequently, the author of this dissertation considered the CDQM Framework (Otto et al., 2011), the associated EDQM Maturity Model (Ofner, Otto et al., 2013), and the more detailed, practitioner-oriented description of the maturity model (EFQM, 2016) as the foundations of his design activities. The structure and content of the DXM is the result of six key decisions made by the author of the dissertation in response to the design requirements (see *Section 6.1*):

1. Conceive data management as a continuous management cycle
2. Translate business capabilities into data management capabilities
3. Explicate a data strategy
4. Develop data management capabilities through governance and technological aspects
5. Emphasize the data lifecycle
6. Demonstrate results of data management in terms of data excellence and business value

Figure 6-3 relates these six design decisions to the design requirements specified in *Section 6.1*. After that, each design decision is described in a separate subsection.

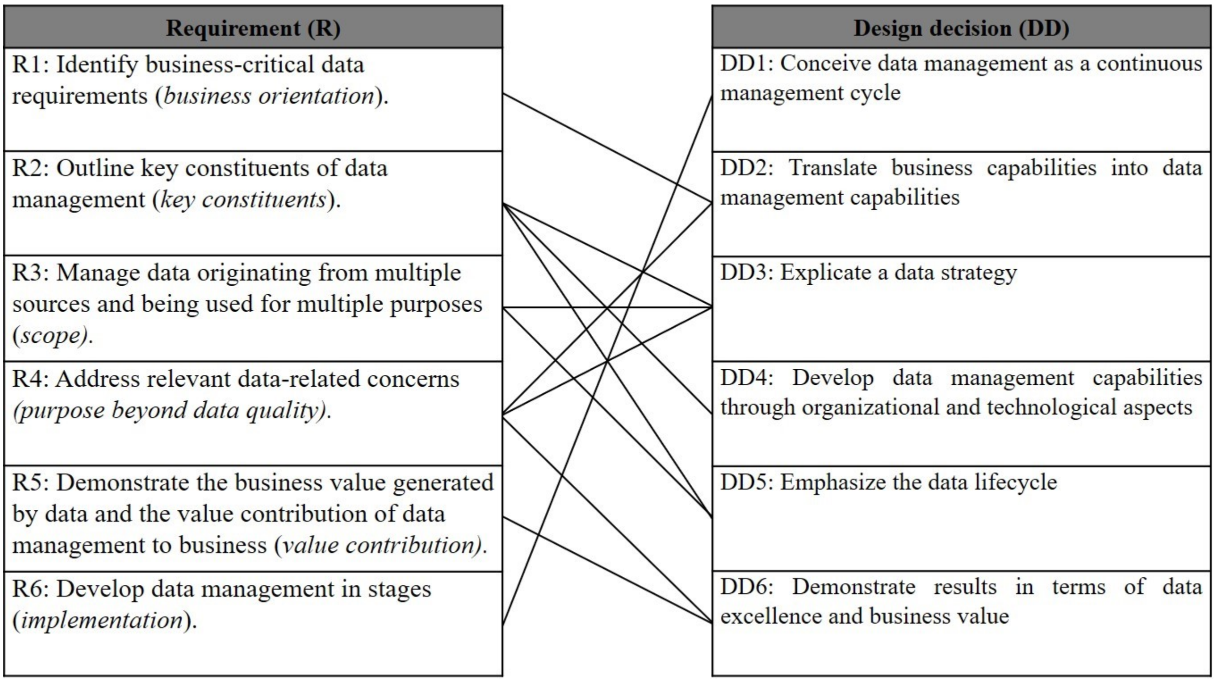


Figure 6-3: Requirements and design decisions

6.3.1 Design Decision 1: Conceive Data Management as a Continuous Management Cycle

To enable data managers in data-driven enterprises to effectively manage data as a strategic resource, one major element of the reference model is Continuous Improvement. Indicating the dynamic nature of strategic data management, this element of the DXM builds on existing approaches for asset management (ISO, 2014a, ISO, 2014b) and performance management (Ferreira & Otley, 2009; Otley, 1999) and adopts the logic of the PDCA cycle (i.e. plan – do – check – act). Continuous Improvement interlinks the different design areas of data management specified by the DXM, which are organized in three categories: Goals define the strategic direction for data management (corresponding with *plan* of the PDCA logic), Enablers facilitate the Goals (i.e. *do*), and Results measure the achievement of the Goals (i.e. *check*). Within this structure, Continuous Improvement establishes a process to adjust the Goals and improve the Enablers based on the Results achieved (i.e. *act*).

Structuring the reference model in accordance with the PDCA cycle logic was proposed by the research team after reviewing EFQM's Excellence Model (2009) and confirmed by representatives of the CC CDQ member companies during the initial plenary discussion (see research activity 1-2 in *Table 4-3*). As a result of the first iteration, the idea of continuous improvement was conceptualized in a design area named "Learning, Creativity & Innovation" (see *Figure A-1* in *Appendix A.3*). This design areas was renamed several times into "Improvement & Alignment" (research activity 1-3, *Figure A-2*), "Innovation & Learning" (research activity 1-5, *Figure A-3*), and "Innovation and Continuous Improvement" (research activity 1-7, *Figure A-4*), up to the point when it received its final name ("Continuous Improvement") in research activity 1-12.

6.3.2 Design Decision 2: Translate Business Capabilities Into Data Management Capabilities

The growing relevance of data for corporate value propositions, together with the possibility to establish different types of data monetization thanks to data analytics (Schüritz et al., 2017; Wixom & Ross, 2017), requires extended data management capabilities. The DXM incorporates an outside-in perspective, indicating that data management is contingent on business objectives. The model's Goal section focuses on capabilities, describing what a company should do to establish strategic data management (Bērziša et al., 2015) and identify business-critical data requirements (Bärenfänger et al., 2016; Bärenfänger & Otto, 2015). By first explicitly documenting data's business-critical role in the form of business capabilities, and then deriving data management capabilities from these business capabilities, data managers are able to directly align their activities to changing business requirements.

Integrating the RBV and a capabilities perspective into the Goals section of the reference model was decided in a focus group discussion during research activity 1-7. The focus group participants rejected a proposal of the research team, which included "Internal & External Stakeholders" and "Corporate Strategies" as influencing factors of the "Data Strategy" design area (see *Figure A-3* in *Appendix A.3*). Instead, the group agreed to link "Data Management Capabilities" to "Business Capabilities" (see *Figure A-4* for the reference model version resulting from this iteration).

6.3.3 Design Decision 3: Explicate a Data Strategy

Considering data as a strategic resource requires a consistent, enterprise-wide management approach. This approach should be documented in the form of a data strategy (DalleMule & Davenport, 2017). Generally speaking, a strategy provides guidance by defining a target state, developing a plan to reach this state, and specifying focus areas and activities (Mintzberg, 1994; Porter, 1996). With regard to data, a strategy is required as the guiding frame to ensure a consistent, enterprise-wide approach concerning the way enterprise data is to be managed and used.

Already at an early stage of the research process (research activity 1-1), the research team proposed a design area labeled “Data/Information Strategy” (see *Figure A-1* in *Appendix A.3*), which was not further discussed in the plenary discussion though. During the subsequent focus group discussion (research activity 1-3), the participants agreed to reflect in the new reference model the six design areas of the CDQM Framework, which already included a design area named “Data Quality Management Strategy”. This design area was then assigned to the Goals section of the reference model and renamed into “Data Strategy” for reflecting a broader scope (see *Figure A-2* for the reference model version resulting from this iteration). The author of the dissertation specified this design area with the help of case studies (research activities 3-2 and 3-9), expert interviews (research activity 3-8), and focus group discussions (research activities 3-3, 3-6, and 3-7).

6.3.4 Design Decision 4: Develop Data Management Capabilities Through Organizational and Technological Aspects

A key lesson learned from applying the Framework for CDQM (i.e. the predecessor of the DXM), which has been confirmed by empirical evidence gained from other research and by the analysis of competing artifacts, is that data management capabilities have a dual nature, comprising both organizational and technological aspects. As data management comprises the planning, provision, organization, use, and disposal of data, it cannot be approached from a purely technological perspective (Jain et al., 1998). In addition to technology-related aspects, data management entails data governance, requires roles and responsibilities (Abraham, Schneider, & Vom Brocke, 2019; Khatri & Brown, 2010; Vilminko-Heikkinen & Pekkola, 2017), depends on processes and methods (Kahn et al., 2003), and uses performance management systems (Pipino et al., 2002; Wang, 1998). Thus, the DXM includes two technological design areas (Data Applications and Data Architecture) and three organizational (or governance-related) design areas (People, Roles, and Responsibilities; Processes and Methods; and Performance Management).

In the first focus group discussion of the design process (research activity 1-3), the participants decided to integrate into the new reference model the six design areas of the CDQM Framework, which – besides a design area referring to strategy (see design decision 3) – included two technological design areas (i.e. “Corporate Data Architecture” and “Corporate Data Applications”) and three organizational design areas (i.e. “Controlling”, “Organization & People”, and “Processes & Methods”) (see *Figure A-2* in *Appendix A.3*).

In the following focus group (research activity 1-5), “Controlling” was renamed into “Performance Monitoring” to circumvent the somewhat biased understanding of the term “Controlling” in the German-speaking countries, from where the majority of the CC CDQ member companies originates. In research activity 1-7, the design area received its final name (“Performance Management”) to reflect the broad scope of practices within this design area, which encompasses more than just performance monitoring (e.g. definition of metrics and measurement activities).

After discussing the “Organization & People” design area in a focus group (research activity 1-7), the research team proposed to change the name into “People, Roles, and Responsibilities” in order to reflect that the design area does not only contain formal organizational aspects (such as the definition of data-related roles and their responsibilities) but also cultural aspects (such as turning “people” into “data citizens”). The proposal was then approved by the participants of the subsequent focus group (research activity 1-9).

While the name of the “Processes & Methods” design area was taken over from the CDQM Framework, focus group participants in research activity 1-7 decided to shorten the terms “Corporate Data Architecture” and “Corporate Data Applications” into “Data Architecture” and “Data Applications”. This name change not only allows for easier (shorter) communication but also mirrors cross-corporate aspects of strategic data management, in which external data (in addition to corporate / internal data) is to be addressed in data architecture and the data application landscape additionally consists of applications outside of a company’s direct control.

6.3.5 Design Decision 5: Emphasize the Data Lifecycle

Data lifecycle management ensures a consistent, enterprise-wide approach to create, maintain, use, and archive data. In the past, there was no transparency regarding the origin of raw data (i.e. the data sources) and the flow of data through an enterprise’s systems and applications. With the proliferation of big data analytics, new data value chains have emerged that require orchestration (Abbasi et al., 2016). By managing the data lifecycle, companies define the processes for creation, acquisition, storage, maintenance, usage, and deletion of data (Redman, 1996).

After deciding to adopt the elements of the CDQM Framework, the focus group participants in research activity 1-5 highlighted the importance of the data lifecycle as a key task to be addressed by a data manager. While data lifecycle processes were part of the CDQM Framework’s “Processes & Methods” design area, the focus group decided to define "Data Lifecycle" as a separate design area in the Enablers section of the reference model.

6.3.6 Design Decision 6: Demonstrate Results in Terms of Data Excellence and Business Value

Being an outcome-oriented capability, strategic data management results in two types of outcomes. First, data management has a direct impact on data quality, defined by the reference model as Data Excellence. The concept of excellence, which originates from TQM (Suarez, Calvo-Mora, & Roldán, 2016), is well-suited for being transferred to the data domain. Extending the limited scope of quality-oriented data management (Batini & Scannapieca, 2006; English, 2003; Wang, 1998; Wang et al., 1998), Data Excellence comprises additional aspects, such as regulatory compliance, data security, or data privacy (Delbaere & Ferreira, 2007; Sadeghi, Wachsmann, & Waidner, 2015).

Second, Data Excellence creates value for business, which is reflected by the Business Value design area of the DXM. Data management's value contribution includes the support of business processes (Reid & Catterall, 2005; Tellkamp et al., 2004; Vermeer, 2000; Zahay & Griffin, 2003), the improvement of decision-making (Orr, 1998; Price & Shanks, 2005; Shankaranarayanan et al., 2003), an increase in enterprise performance (Joshi & Rai, 2000; Sheng & Mykytyn, 2002), and the creation of novel forms of value proposition and data-driven innovation (Schüritz et al., 2017).

While the basic structure of the reference model (made up of Goals, Enablers, and Results) was confirmed in research activity 1-2, the elements of the Results section were discussed and changed in multiple iterations. A first feedback from focus group participants in research activity 1-3 suggested that the four sub-criteria of the EDQM maturity model's Results section (i.e. "Customer Results", "Employee Results", "Compliance Results", and "Business Performance"; see *Figure A-1* in *Appendix A.3*) that were initially proposed to be used for the reference model were too complex and difficult to assess and communicate. Consequently, the research team proposed only one Results design area, named "Data Management Results". In the following focus group (research activity 1-5), the participants agreed to distinguish between "Business Impact" and "Data Management Impact" in the Results section (see *Figure A-3* in *Appendix A.3*).

Participants of another focus group (research activity 1-7) then decided to change "Data Management Impact" into "Data Excellence". This term was selected as it conveys the notion of continuous improvement through the ambition of organizations to become "excellent". Furthermore, the idea was to coin an entirely new term that was not known of being used by anyone else so far, and which could be used as an umbrella term that spans several areas organizations need to strive for continuous improvement in, such as data compliance, data privacy, data security, and – of course – data quality. Finally, focus group participants in research activity 1-9 agreed to change “Business Impact” into “Business Value”, as “value” was perceived to have a more positive connotation compared to “impact”.

6.4 Role Model

Managing data as a strategic resource requires different data management related roles within an enterprise. Consequently, a capability reference model for strategic data management must allow for being used by each of these roles. The author of this dissertation specified these roles with the help of a role model for data catalogs, which was developed by the CC CDQ and Fraunhofer Institute for Software and Systems Engineering (ISST) (Korte et al., 2018). He did so, as data catalogs – besides enterprise analytics platforms – are a key data application for data-driven organizations (see *Subsection 2.1.2*), and as users of data catalogs typically come from various corporate functions, indicating the broad relevance of data when it comes to striving for strategic data management. The role model consists of eight user groups (Korte et al., 2018, pp. 16–17): (1) Chief Data Officer / Data Manager, (2) Data Citizen, (3) Data Owner, (4) Data Analyst, (5) Data Protection Officer, (6) Data Steward, (7) Data Architect, and (8) Solution Architect. To reflect the different decisions to be taken on a strategic level (i.e. board) and a governance level, the author of the dissertation split one role (Chief Data Officer / Data Manager) into two roles (Data Management Sponsor and Data Manager), which are more suitable for the purpose of the reference model. *Table 6-3* lists the nine data management roles related to the DXM, introduces each role with an illustrative statement, defines the role’s responsibilities, presents justificatory knowledge, and outlines how the role can benefit from applying the DXM.

Table 6-3: Role description of DXM users

Role	Description
Data Management Sponsor Similar roles are Chief Data Officer (CDO) or Chief Information Officer (CIO)	**Illustrative statement:** “As a Data Management Sponsor, I ensure funding and sponsorship of data management on board level.” **Responsibilities:** The Data Management Sponsor is the company’s strategic head of data management. He/she steers and promotes data management activities on the executive level and ensures financial sponsorship.

Role	Description
	Justificatory knowledge: (DAMA, 2017, p. 76; Griffin, 2008, p. 28; Korte et al., 2018, p. 16; Leimeister, 2015, p. 5; Weber, 2009, p. 107; Xu et al., 2016) **Benefits from using the reference model:** The Data Management Sponsor can use the reference model to demonstrate the interrelations between data management and business to the other members of the executive board (i.e. Business Capabilities and Data Management Capabilities in the Goals section of the DXM). Furthermore, the DXM allows the Data Management Sponsor to create transparency regarding the value contribution of data management to the business (i.e. Business Value in the Results section). In addition, the DXM helps the Data Management Sponsor communicate the data strategy both within the enterprise and externally.
Data Manager Similar roles are Head of Data Management, Data Governor, Corporate Data Steward, or Strategic Data Steward.	**Illustrative statement:** "As a Data Manager, I orchestrate all data management related activities of the company." **Responsibilities:** The Data Manager is the leading role in all data management activities, as he/she specifies and implements the data strategy. By doing so, he/she designs the organizational and technological design areas of data management (i.e. People, Roles and Responsibilities; Processes and Methods; Data Lifecycle; Data Applications; Data Architecture; and Performance Management). Furthermore, he/she evaluates the Results of data management (i.e. Data Excellence and Business Value) and initiates measures for improvement. **Justificatory knowledge:** (DAMA, 2017, pp. 76–77; Korte et al., 2018, p. 16; Loshin, 2009; Weber, 2009, p. 107) **Benefits from using the reference model:** The DXM allows the Data Manager to specify the data strategy (i.e. the design areas of the Goals section of the DXM), operationalize the data strategy (i.e. the design areas of the Enablers section), evaluate the outcomes of data management (i.e. the design areas of the Results section), and continuously improve data management. Furthermore, he/she can use the reference model to communicate data management to the business and assess the maturity of data management activities.
Data Citizen Similar roles are Data User, Data Consumer, or Business User.	**Illustrative statement:** "As a Data Citizen, I rely on the data's excellence to use it most effectively in my day-to-day business activities." **Responsibilities:** The Data Citizen consumes data products and receives (basic) data management services. He/she works in a business department (i.e. is not assigned to the data management organization) and specifies Data Excellence requirements. **Justificatory knowledge:** (Korte et al., 2018, p. 16; Loshin, 2009, p. 28) **Benefits from using the reference model:** The DXM helps the Data Citizen understand the design areas, the basic terms, and definitions of data management.
Data Owner Similar roles are Data Definition Owner, Data Structure Owner.	**Illustrative statement:** "As a Data Owner, I define the structure and content of the data I am responsible for, and I interact with the consumers of "my" data to meet their requirements." **Responsibilities:** The Data Owner defines his/her data (i.e. detailing the metadata per object and attribute). By doing so, he/she interacts with the data users, gathers their Data Excellence requirements, and makes sure

Role	Description
	their expectations are met. The Data Owner also defines data access rights and allows or prohibits certain data usage scenarios. **Justificatory knowledge:** (Korte et al., 2018, p. 16; Redman, 2001, p. 184; Reichert, 2015, p. 93) **Benefits from using the reference model:** The DXM helps the Data Owner specify the data lifecycle of his/her data, understand the interrelations of his/her data with other data objects, and recognize the storage, distribution, and flow of his/her data (i.e. the Data Architecture).
Data Analyst Other (more specific) roles that can be subsumed under Data Analyst are Data Scientist, Data Expert, or Business Intelligence (BI) Analyst.	**Illustrative statement:** "As a Data Analyst, I develop data analytics use cases by combining data from multiple internal and external sources." **Responsibilities:** The Data Analyst collects and transforms (i.e. integrates and aggregates) data from multiple sources to create data (analytics) products. To do so, he/she defines data supply chains and applies analytics techniques. **Justificatory knowledge:** (Davenport & Patil, 2012; Korte et al., 2018, p. 16) **Benefits from using the reference model:** The DXM allows the Data Analyst to derive the portfolio of data products and data management services from the business requirements, and to define the necessary data management capabilities (i.e. Business Capabilities and Data Management Capabilities in the Goals section of the DXM). Furthermore, he/she can use the reference model to design the data supply chain and define the data sources and technologies required for creating data products (i.e. Data Lifecycle; Data Applications, and Data Architecture in the Enablers section of the DXM).
Data Protection Officer Similar roles are Data Security Officer or Data Compliance Officer.	**Illustrative statement:** "As a Data Protection Officer, I define access to data and make sure it is used for the intended purpose only." **Responsibilities:** The Data Protection Officer identifies the data and datasets that must be specifically protected from unauthorized access. He/she flags the critical attributes, and defines and controls the mechanisms to protect these datasets. **Justificatory knowledge:** (Korte et al., 2018, p. 17) **Benefits from using the reference model:** The DXM helps the Data Protection Officer identify the relevant stakeholders (i.e. People, Roles, and Responsibilities), understand data-related processes, define protection mechanisms (i.e. Processes and Methods, Data Lifecycle), and document which data is to be protected (i.e. Data Architecture).
Data Steward Similar roles are Data Quality Manager or Data Content Steward.	**Illustrative statement:** "As a Data Steward, I define how data is created, maintained, and used in accordance with Data Owners' requirements." **Responsibilities:** The Data Steward documents business rules based on Data Owners' requirements. He/she specifies data with the help of metadata, defines data lifecycle processes, and defines and controls data excellence metrics. **Justificatory knowledge:** (Bärenfänger, 2017, p. 179; DAMA, 2017, p. 77; Korte et al., 2018, p. 17; Weber, 2009, p. 107) **Benefits from using the reference model:** The DXM helps the Data Steward understand the link between data management goals (i.e. the

Role	Description
	Data Excellence dimensions targeted and the goals of the Data Strategy) and business goals (i.e. Data Excellence and Business Value in the Goals section of the DXM). Furthermore, the DXM allows the Data Steward to review and improve the processes and methods of data management and performance management measures, and define the required data lifecycle activities (i.e. Processes and Methods, Data Lifecycle, Performance Management in the Enablers section of the DXM).
Data Architect Similar roles are Enterprise Architect or Data Modelling Expert.	**Illustrative statement:** "As a Data Architect, I define how data is stored in and consumed by applications, and I develop and maintain data models for different stakeholders in business, data management, and IT." **Responsibilities:** The Data Architect creates and maintains data documentations and data models on a conceptual, canonical, and physical level. He/she defines and monitors how data is stored and used in the data application landscape. **Justificatory knowledge:** (Korte et al., 2018, p. 17; Schmidt, p. 169) **Benefits from using the reference model:** The DXM helps the Data Architect translate business requirements into data models. Using the DXM, he/she is enabled to understand the interrelations between his/her domain (i.e. Data Architecture) and related design areas, such as Data Applications or Data Lifecycle.
Solution Architect Similar roles are Data Engineer, IT/Data Product Engineer, or Technical Data Steward.	**Illustrative statement:** "As a Solution Architect, I implement the concepts and methods defined by the Data Steward and the Data Architect in our applications, and I support the Data Analyst in discovering and mapping data schemas." **Responsibilities:** The Solution Architect is responsible for the technical implementation of business specifications and data management specifications. He/she discovers application-specific data schemas and maps data schemas across applications. **Justificatory knowledge:** (DAMA, 2017, p. 77; Korte et al., 2018, p. 17; Wang et al., 1998, p. 98; Weber, 2009, p. 107) **Benefits from using the reference model:** The DXM helps the Solution Architect understand business and data management requirements. Using the DXM, he/she is able to review Data Management Capabilities, and to specify the functions required from data applications to provide and support these capabilities (i.e. Data Applications). In addition, he/she is enabled to review Data Excellence objectives in order to provide adequate IT solutions.

6.5 Reference Model Description

In the following twelve subsections, the design areas of the DXM are specified and illustrated in detail. Each subsection is structured as follows:

- First, the author provides a definition[38], gives a short description of the design area, and presents the overall ambition towards excellence.

- Second, the author introduces success criteria for the design area. To ensure the success criteria actually materialize in concrete actions, a set of recommended practices is given. For the purpose of further illustration of the success criteria, the author includes references to the case studies conducted in relation to this dissertation.

- Third, the success criteria are supported with theory-based and/or empirical evidences as justificatory knowledge. Each success criterion evolved, starting with the *success criterion's definition* (i.e. identifying and defining the criterion and its practices), to *adoption* (i.e. collecting further evidences and maintaining the criterion), to *extension* (i.e. broadening the scope) and *modification* (i.e. improving, changing, or correcting). The author indicates whether the success factor was newly introduced in the design process of the DXM (*definition*) or taken over from its predecessors, the CDQM Framework and/or EDQM Maturity Model, without changes (*adoption*), with a broader scope (*extension*), or with significant adjustments (*modification*).

- Fourth and final, the author presents a set of documents that typically result from the activities performed in relation to the design area, including selected artifacts (i.e. models or methods known from literature, mainly from the CC CDQ context) that can be used to support the implementation of the recommended practices.

6.5.1 Business Capabilities

Definition: Business capabilities define a set of data-based skills, routines, and resources a company needs to have in order to achieve its business goals through data monetization.

Business capabilities describe what a company does, or should do (Bērziša et al., 2015), in order to achieve certain business goals. The Business Capability design area specifies what data-related business capabilities are required, which of these are already in place to some extent and need to be enhanced, and which ones need to be established from scratch. Data-related business capabilities describe how a company monetizes data in terms of operational excellence (including data-driven insights and decision-making),

[38] The Glossary in the appendix comprises all definitions of the DXM design areas and constructs.

new or enhanced business models (Schüritz et al., 2017; Wixom & Ross, 2017), and risk reduction and enhanced compliance (Fadler & Legner, 2019). To define the relevant business capabilities, a detailed understanding of the business and its requirements with regard to data is a prerequisite (Azevedo et al., 2015). Furthermore, this design area details how a data-driven working mode is to be established as corporate culture plays a key role in operationalization of business capabilities (Alexander & Lyytinen, 2017; Collis, 1994).

Excellence statement 1: Excellent organizations recognize the importance of data as a prerequisite to respond to business drivers and achieve business objectives. They exploit data for operational excellence, explore data-related business opportunities, and promote a data-driven working mode.

Three success criteria define excellent organizations in the Business Capabilities design area (see *Table 6-4*):

a. Given data's business criticality, excellent organizations review and document requirements on data from a strategic (top-down) and an operational (bottom-up) perspective. Strategically, they consider the company's external environment, which is defined by e.g. customer expectations, market regulation, or market developments as a result of competitors' activities. Also, the companies' business strategy (i.e. its vision and mission) as well as their strategic business goals are understood by data management. Operationally, excellent organizations review the impact of data on business, derive and document data-related requirements.

> *Tobacco company PMI (see Subsection 5.3.1) designed its data management activities based on the requirements of the transformation of its business, which were derived in a top-down approach.*

> *Pharmaceutical company Bayer (see Subsection 5.4.1) started its data strategy development process by analyzing business stakeholders, their activities, and their needs regarding data (see Figure 5-3 on p. 101). The company exemplified how business capabilities can be derived and documented in a bottom-up approach.*

b. Excellent organizations exploit data for operational excellence (in business processes and decision making), assess and implement new ways of exploring data for top-line growth (through new or enhanced business models), bottom-line growth (through data-driven insights), risk reduction, and enhanced compliance. To do so, they identify and document data-related use cases.

> *Automotive supplier Bosch (see Subsection 5.4.5) gathered and reviewed use cases exploiting and using data to develop its data strategy.*

c. Finally, excellent organizations have a data-driven working mode, which makes data an inherent aspect of all business actions. Their employees are data literate and handle data in a responsible way.

> *PMI's data initiative (see Subsection 5.3.1) had the slogan "every decision, every day - with data" to underline that data is an inherent aspect of decision making; and it aimed at "disrupting and transforming the way people at PMI think about and use data".*

> *Software provider SAP (see Subsection 5.3.2) promoted the concept of data citizenship to create a data-driven working mode.*

Table 6-4: Success criteria, recommended practices, justificatory knowledge, and evolution of the "Business Capabilities" design area

Success criteria	Recommended practices	Justificatory knowledge	Evolution
1a. Business requirements towards data are continuously analyzed and documented.	− Review and translate the corporate strategy, further strategies (e.g. digitalization strategy, functional strategies), and business drivers for requirements on data (in a top-down approach) − Determine and analyze the role of data in business models and operational excellence (in a bottom-up approach) − Document and communicate business requirements on data	EDQM Maturity Model criterion 1a (Ofner, Otto et al., 2013, p. 14) Business relationship management capability (Bärenfänger, 2017, p. 156) PMI case, Bayer 1 case DXM Maturity Assessment criterion 1.1 (see *Appendix C.1*)	Modification
1b. New ways of using data in business models, business processes, and decision making are explored.	− Assess and implement new ways of exploiting and using data in terms of o operational excellence, o new or enhanced business models, and o risk reduction and enhanced compliance	EDQM Maturity Model criterion 1a (Ofner, Otto et al., 2013, p. 14) Data innovation R&D as a Forrester data management capability (Hopkins et al., 2018) Data is monetized in business models or value	Modification

Success criteria	Recommended practices	Justificatory knowledge	Evolution
	− Identify and document data-related use cases	propositions, business processes, decision-making, and insights generation (Schüritz et al., 2017; Wixom & Ross, 2017). Bosch case DXM Maturity Assessment criterion 1.6 (see *Appendix C.1*)	
1c. The organization has a data-driven working mode.	− Make data an inherent aspect of all business processes, decisions, and activities − Ensure that employees are data-literate and treat data responsibly	EDQM Maturity Model criteria 3b and 3c (Ofner, Otto et al., 2013, p. 14) PMI case, SAP 1 case DXM Maturity Assessment criterion 2.3 (see *Appendix C.1*)	Modification

Activities in the Business Capabilities design area result in documentations of key data domains and objects, business requirements on data, and data-related use cases (see *Table 6-5*). These key results provide the basic understanding of business needs and objectives with regards to data, which builds the foundation for managing data as a strategic resource.

Table 6-5: Key result documents and supporting artifacts of the "Business Capabilities" design area

Key result documents	Supporting artifacts
− Key data domains and objects − Business requirements on data (from e.g. business strategy and goals, initiatives, external business contingencies) − Documentations of data-related use cases and the required data-related business capabilities	Business capability identification method (Bärenfänger, 2017, p. 162) Capability modeling (Azevedo et al., 2015)

6.5.2 Data Management Capabilities

Definition: Data management capabilities define a set of skills, routines, and resources a company needs to have in order to accomplish data excellence that results in business value.

Given the understanding of data management as a dynamic capability to successfully manage and deploy data (Otto, 2012a), data management is contingent on business objectives (Jain et al., 1998). This design area derives the necessary data management capabilities from the business requirements identified for the previous design area. In analogy to the Business Capabilities design area, data management should draw up a data management capability map, specifying what data management capabilities are required, which of these are already in place to some extent and need to be enhanced, and which ones need to be established from scratch.

Data management services and excellent data result from combining several data management capabilities and/or data sets. Data management services are generated by combining several data management capabilities and resources to deliver services such as methodological support for developing data quality metrics or business user support in mass changes of data. "Excellent data" refers to the provision of (raw) data sets made available in high quality and in compliance with regulations, data protection, and data security requirements.

Excellence statement 2: Excellent organizations manage data as a strategic resource, provide data management services and excellent data to support data exploitation and exploration.

Organizations being excellent in the Data Management Capabilities design area consider two success criteria (see *Table 6-6*):

a. Excellent organizations analyze business requirements to provide data management services and excellent data fulfilling the needs of the business (Otto, 2011a).

> *Bayer (see Subsection 5.4.1) derived data management capabilities from business capabilities in order to define master data management services.*

> *Schaeffler (see Subsection 5.4.4) applied the same approach for developing a management approach for sensor data.*

> *PMI (see Subsection 5.3.1) developed a demand-gathering process and use case delivery model to analyze potential data sciences use cases and provide the required data products.*

b. In accordance with the portfolio management concept of IT services (Peppard, 2003; Zarnekow, 2004), excellent organization cluster data provision and data management services in a portfolio offered by data management. They manage this portfolio by defining ownership and responsibilities and monitor the usage of each individual product and service. Furthermore, they regularly review the entire portfolio in order to adapt the offering to changing business demands (Bärenfänger, 2017).

> *Bosch (see Subsection 5.4.5) identified and documented data-related use cases and derived the data management capabilities required for each use case. Based on this analysis, the data strategy did not only define the key enabling design areas of data management but also outlined the portfolio of data management services (e.g. "We establish standards for data usage […]"),the "Data Management Offices provide training […]").*

> *PMI's data products (see Subsection 5.3.1) had a dedicated data product owner. The data product's usage and value contribution were constantly monitored and reported by the owner.*

Table 6-6: Success criteria, recommended practices, justificatory knowledge, and evolution of the "Data Management Capabilities" design area

Success criteria	Recommended practices	Justificatory knowledge	Evolution
2a. Data provision and data management services are defined to support business capabilities.	− Systematically analyze business capabilities to define the required data provision and data management services − Identify which data provision and data management services are already in place but need to be enhanced, and what need to be established from scratch − Identify users and relevant stakeholders of data and data management services, and engage with them to meet their needs and expectations	Business relationship management capability A 1.3 (Bärenfänger, 2017, p. 156) Data management is a dynamic capability allowing companies to deploy data resources (Otto, 2012a). Data management is contingent on business objectives and business capabilities (Jain et al., 1998). Bayer 1 case, Schaeffler case, PMI case DXM Maturity Assessment criterion 1.2 (see *Appendix C.1*)	Definition
2b. A portfolio of data provision	− Prioritize data provision and data management services	Organizations in the digital economy define an	Definition

Success criteria	Recommended practices	Justificatory knowledge	Evolution
and data management services is developed and maintained.	based on business needs and required (resource) investments – Manage the lifecycle of the data provision and data management services portfolio – Assign ownership of each data product and data management service	information service portfolio: IS portfolio management capability A 2.1 and IS portfolio integration capability A 2.2 (Bärenfänger, 2017, p. 156). Bosch case, PMI case DXM Maturity Assessment criterion 1.4 (see *Appendix C.1*)	

Activities in the Data Management Capabilities design area result in a portfolio and in documentations of data provision and data management services (see *Table 6-7*).

Table 6-7: Key result documents and supporting artifacts of the "Data Management Capabilities" design area

Key result documents	Supporting artifacts
– Portfolio of data provision and data management services – Documentation of data provision and data management services	Information service capabilities design method (Bärenfänger, 2017, p. 156), Reference method for information service development (Bärenfänger, 2017, 210 et seqq.) Portfolio management of IT services (Peppard, 2003; Zarnekow, 2004)

6.5.3 Data Strategy

Definition: A data strategy defines a target state in terms of how data should be managed and used across the entire company, and it develops a plan for reaching this target state.

As data is a strategic resource concerning various internal and external stakeholder groups, data management must be considered as a consistent, enterprise-wide endeavor, which must be carefully planned, conducted, and monitored at a strategic level. A data strategy documents this approach. The research domain of strategic management has provided several definitions of a strategy, which can be transferred to the data management domain[39].

In line with these general definitions for corporate strategies, a data strategy defines a target state in terms of how data should be managed and used across the entire company, and it develops a plan for reaching this target state (Adelman, Moss, & Abai, 2005; Lee et al., 2006; Pierce, 2004; Vesely, 1990). DalleMule & Davenport (2017) highlight the relevance of a data strategy for any company, including both defensive and offensive aspects of data management. While defensive elements aim at establishing control over data (by ensuring data security, privacy, integrity, quality, compliance, and governance), offensive elements address data usage in order to gain competitive edge and increase profitability. Defensive aspects of a data strategy relate to data extraction, standardization, storage, and access with the aim for a SSOT, while offensive aspects include data analytics, visualization, transformation, and enrichment with the aim for MVOT.

Excellence statement 3: Excellent organizations define and implement a data strategy, which specifies how to build a data foundation and how to use data in order to generate business value.

Three success criteria are addressed by organizations, which are excellent with regards to the Data Strategy design area (see *Table 6-8*):

a. Excellent organization develop their data strategy based on business requirements. Their data strategy includes a link to the business, a vision and mission statement, data-related objectives, a definition of the boundaries and scope, as well as guiding principles (see the Data Strategy Canvas on p. 163).

> *Bayer's master data management strategy (see Subsection 5.4.1) presented data-related business requirements, a vision and a mission statement, a*

[39] Chandler (1962) defines "strategy" as "the basic long-term goals of an enterprise, and the adoption of courses of action and the allocation of resources necessary for carrying out these goals" (p. 13). According to Andrews (1997), a strategy describes "the pattern of decisions in a company that determines and reveals its objectives, purposes, or goals, produces the principal policies and plans for achieving those goals, and defines […] contribution it intends to make to its shareholders, employees, customers, and communities" (p. 52). Porter (1980) conceives of a strategy as a "broad formula for how a business is going to compete, what its goals should be, and what policies will be needed to carry out those goals" (p. xvi). Rumelt (2011) states that "a good strategy has, at a minimum, three essential components: a *diagnosis* of the situation, the choice of an overall *guiding policy*, and the design of *coherent action*" (p. 268).

> *definition of the strategy's scope and goals, and a roadmap for strategy implementation.*

> *The Bosch data strategy (see Subsection 5.4.5) addressed all areas of the Data Strategy Canvas in a detailed manner and considered both aspects of data strategies: data foundation (defense) and data monetization (offense).*

> *Similarly, PMI's data strategy (see Subsection 5.3.1) included both defensive and offensive elements.*

> *SBB (see Subsection 5.4.3) regularly updated its data strategy: a first version was issued for the period between 2013 and 2016; the current version spans from 2017 to 2020; and SBB already prepared the development of a data strategy for 2021 and beyond.*

b. Top managers of excellent organizations are personally involved in defining and implementing the data strategy. They assume ownership of the data strategy, provide support and sponsorship on board level for data-related activities (Griffin, 2008; Xu et al., 2016), and underline the importance of data through explicit actions, decisions, and supportive statements.

> *PMI (see Subsection 5.3.1) established the role of the Chief Analytics and Data Officer, who led the Enterprise Data & Analytics unit, was responsible for and owned the data strategy, and promoted all data-related topics on executive board level.*

c. Excellent organization operationalize their data strategies with an implementation roadmap, which defines the priorities and sequence of actions. They secure availability of the resources required and ensure that relevant aspects of the data strategy are embedded in the organization's structure, corporate policies, standards, and guidelines.

> *The implementation layer of Bosch's data strategy (see Subsection 5.4.5) outlined the strategic roadmap following a use case driven approach, starting with the use cases generating the highest value, and detailed the planned actions per activity block.*

PMI (see Subsection 5.3.1) defined a staged approach for implementing its data strategy. During the first phase, the priority was on establishing data governance and delivering first results from data science. In this phase, EAD focused on four building blocks, namely Data Management, Data Architecture, Data Science, and the Program Management Office. Further elements of the strategy (such as data quality management) were of less importance and planned for the second phase.

To operationalize its master data management strategy (see Subsection 5.4.1) in daily business operations, Bayer issued a new version of its corporate directive on master data management (see Subsection 5.4.2).

Table 6-8: Success criteria, recommended practices, justificatory knowledge, and evolution of the "Data Strategy" design area

Success criteria	Recommended practices	Justificatory knowledge	Evolution
3a. Vision and mission for managing and using data are developed and communicated.	− Define/update and communicate the target picture for managing and using data (vision) and the data organization's role (mission) − Define/update and communicate concrete data-related objectives − Define/update and communicate boundaries and scope of the data strategy − Define/update and communicate principles guiding and shaping the future culture with regard to data management and usage	EDQM Maturity Model criterion 1a (Ofner, Otto et al., 2013, p. 14) A data (management) strategy is an important aspect of data management (DAMA, 2017, pp. 32–48). Data management strategy as key component (EDM Council, 2018) Mission & vision, goals & objectives, guiding principles, success measures, action plans as data quality capabilities (GS1, 2010) Vision and strategy as enterprise information management building blocks (Gartner, 2014) Bayer 1 case, Bosch case, PMI case, SBB 1 case DXM Maturity Assessment criterion 1.3 (see *Appendix C.1*)	Extension
3b. Top managers are personally involved in defining	− Ensure all data-related activities are aligned with top management's goals	EDQM Maturity Model criterion 1b (Ofner, Otto et al., 2013, p. 14)	Modification

Success criteria	Recommended practices	Justificatory knowledge	Evolution
and implementing the data strategy.	− Ensure and maintain support of and sponsorship from board members − Define and ensure clear ownership of data strategy and every associated data activity	"Leadership is committed to data quality principles, relies on data for decision making, and creates a culture of data usage." (PIC, 2016, p. 9). Executive sponsorship and accountable leadership as data quality capabilities (GS1, 2010) PMI case DXM Maturity Assessment criterion 2.2 (see *Appendix C.1*)	
3c. The data strategy is integrated in the corporate management system.	− Develop a data strategy implementation roadmap (including definition and review of priorities in data management and other related activities) − Secure the resources required to implement the data strategy − Ensure that relevant aspects of the data strategy are incorporated in corporate policies, standards, and guidelines − Define the organization's operating model to support the development, ownership, and implementation of the data strategy	EDQM Maturity Model criterion 1b (Ofner, Otto et al., 2013, p. 14) Availability of basic resources (Gupta & George, 2016, p. 1051) Bosch case, PMI case, Bayer 2 case DXM Maturity Assessment criterion 1.5 (see *Appendix C.1*)	Modification

Activities in the Data Strategy design area result in a data strategy document, a code of conduct, and an implementation roadmap (see Table 6-9).

Table 6-9: Key result documents and supporting artifacts of the "Data Strategy" design area

Key result documents	Supporting artifacts
– Data strategy statement and document (specifying need for action, vision, mission, scope, business value, and key capabilities) – Code of conduct (guiding and shaping the future culture with regard to data management and usage) – Data strategy implementation roadmap	Data Strategy Canvas (see below); Strategy Design Method for CDQM (Falge, 2015; Falge, Otto, & Österle, 2013); Strategic planning in excellence management systems (Suarez et al., 2016)

Data Strategy Canvas

As a response to the changing role of data (in the era of strategic data management), member companies of the CC CDQ faced the challenge of revising their data strategies and wondered which elements a data strategy should contain. Following an initial focus group discussion in the summer of 2018 (research activity 3-3), a series of 23 expert interviews conducted in the autumn of 2019 (research activity 3-8), and additional focus group discussions taking place throughout 2019 (research activities 3-6 and 3-7), a CC CDQ research team (which the author of this dissertation was a member of) developed the Data Strategy Canvas (*Figure 6-4*).

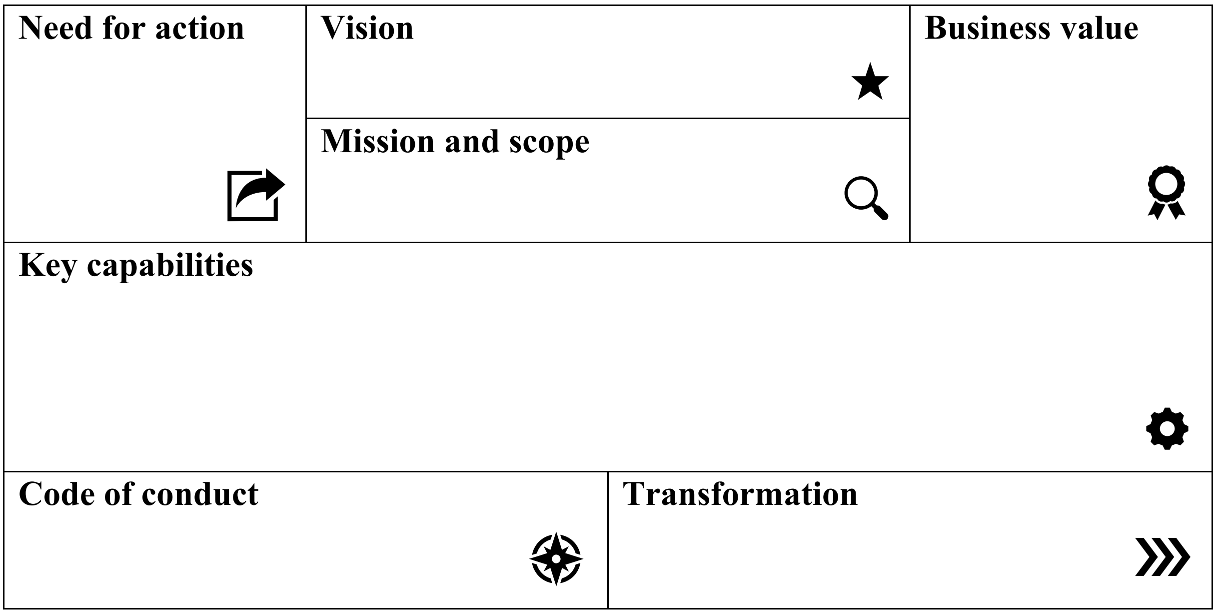


Figure 6-4: The Data Strategy Canvas

The Data Strategy Canvas specifies the following elements to be taken into account by a data strategy:

- **Need for action:** Managing data as a strategic resource requires that data management understands the business requirements. Consequently, a data strategy motivates data-driven activities and reflects the current level of data management maturity. This element of the data strategy answers two questions: *Where do we stand? Why do we have to change?*
- **Vision:** A data strategy includes a vision, which – as a desired state to be achieved in the future – defines the aspiration expressed by the data strategy. This element answers the question: *What role should data play in our company in the future?*
- **Mission statement and scope:** A data strategy includes a mission statement (i.e. a view of what the organization should do, and why and how this should be done). Based on this mission statement, a data strategy specifies the scope to be covered (i.e. which data types and domains should be included). The question answered by this element is: *What should be the purpose and scope of our data management initiatives and organization?*
- **Business value:** A data strategy is a communication tool promoting the business-criticality of data. This element answers the question: *What is the value contribution of data to the business?*
- **Key capabilities:** To accomplish the vision and generate the business value desired, the data strategy defines the key capabilities required. This element answers the question: *Which organizational and technical capabilities need to be improved, and what needs to be built up from scratch?*
- **Code of conduct:** A data strategy defines the guiding principles for managing and using data, thereby providing guidance in daily operations by reducing the number of options to choose from. The code of conduct describes the future mindset and culture related to data for both internal stakeholders (i.e. employees) and external stakeholders (i.e. customers and partners). The questions answered by this element is: *What should be the guiding principles of data management in the future?*
- **Transformation:** For being operationalized, a data strategy defines a roadmap specifying concrete initiatives as well as a resource plan to ensure availability of resources. This element of the data strategy answers two questions: *How should*

the data strategy be implemented? What resources should be used to implement the data strategy, and how should they be deployed?

6.5.4 People, Roles, and Responsibilities

Definition: People, roles, and responsibilities define the culture, organization, roles, boards, and interactions for strategic data management.

Data is generated, managed, and used in many different parts of an organization. A formal description of the data management organization supports the orchestration and alignment of enterprise-wide data management activities. This is of particular importance as data management organizations are typically federated (or virtual), i.e. employees remain in their original functions and chain of command, but in addition are responsible for certain data-related activities. Consequently, data can only be managed consistently if ownership and responsibilities are assigned, and trained (Abraham et al., 2019; Khatri & Brown, 2010; Vilminko-Heikkinen & Pekkola, 2017). A data management organization manifests itself in a concrete organizational design with defined roles and boards, patterns of interaction, and an appropriate, data-driven corporate culture. For example, a Data Management Sponsor is responsible for ensuring funding and promoting data management on a top management level (Weber, 2009, p. 107 et seq.), while a Data Owner is responsible and accountable for providing unambiguous definitions of ”her/his” data, collecting the requirements of business processes using their data, and defining measures of data excellence to ensure business requirements are met.

Regarding roles and boards, different concepts can be observed on a strategic, governance, and operational level. In more defensive approaches, typical roles are Data Manager, Data Owner, Data Steward, Data Architect, Solution Architect, and Data Protection Officer (Otto & Österle, 2015; Weber et al., 2009b, p. 12). More offensive approaches include the CDO, Data Citizen, and Data Analyst as additional roles (Davenport & Patil, 2012; Korte et al., 2018; Xu et al., 2016) (see *Section 6.4*).

Excellence statement 4: Excellent organizations maintain a data-driven culture, they develop, manage, and leverage their employees’ potential regarding data on an individual, team, and organizational level.

Organizations being excellent in the People, Roles, and Responsibilities design area address four success criteria (see *Table 6-10*):

a. Excellent organizations define and update data-related roles, responsibilities, and decision-making rights. As data-related responsibilities are executed mainly in business activities, excellent organizations embed data-related duties and rights in existing structures and roles.

> *The Bosch data strategy (see Subsection 5.4.5) defined and specified six data-related roles: Chief Data Officer, Data Owner, Data Possessor, Data Steward, Data Custodian, and Data Consumer.*

> *Similarly, Bayer's corporate master data management directive (see Subsection 5.4.2) introduced seven data-related roles and four boards and defined their responsibilities and decision rights: Global Process Owner – Master Data, Functional Data Steward, Operational Data Steward, Master Data Coordinator, IT Functional Leader, IT Design and Build, and IT Execute, as well as Master Data Governance Council, Master Data Governance Sub-Council, Master Data Domain Board, and Master Data Operational Community.*

> *In its governance policy, SBB (see Subsection 5.4.3) integrated data-related responsibilities into existing role descriptions. For instance, ownership of a data object was assigned to the owner of the process which creates the data object.*

> *tesa (see Subsection 5.5.2) made use of existing team meeting structures (such as department meetings or management meetings) to effectively position data management as a relevant topic*

b. To coordinate and facilitate data-related activities in a federated data community, excellent organizations establish a central data management organization, which develops, maintains, promotes, and controls corporate policies, standards, and guidelines for data management. In addition, by defining patterns of interaction it is possible to outline how the central data management organization and federated roles are to interact with each other in terms of reporting lines, communication flows, and locus of decision-making.

> *The Bosch data strategy (see Subsection 5.4.5) outlined principles for an operating model defining the Central Enterprise Architecture team as the*

> *coordinating unit and explaining the patterns of interaction between data-related roles. These role descriptions were embedded in existing roles profiles in the organization.*

> *Schaeffler's Corporate Data Management team (see Subsection 5.4.4) centrally defined methods and standards, which are to be adopted by the (local/functional) managers of all data domains.*

c. Excellent organizations establish and manage a data-driven culture, in which all employees understand the importance of data (Chen et al., 2012a; LaValle et al., 2011). Their employees are aware that they – as data citizens – have duties and rights.

> *tesa's Corporate Data Management (see Subsection 5.5.2) established a data community by organizing a bi-weekly "master data table" during lunch break in order to give all interested employees an opportunity to informally exchange. Furthermore, tesa employees were informed about the Corporate Data Management's team, mission, and service offering in a company magazine.*

> *Software provider SAP (see Subsection 5.3.2) promoted the concept of data citizenship via various communication channels to create a data-driven culture.*

d. Excellent organizations empower their employees to assume responsibility for data. For each data-related role, they define the knowledge, skills, and information required, run a training and education program (Khatri & Brown, 2010), plan and manage the career development.

> *The Bosch data strategy (see Subsection 5.4.5) defined principles for a training concept as a key deliverable of the People, Roles, and Responsibilities design area. This concept aimed at equipping role owners with the right skills and competencies to fulfil their responsibilities.*

> *The CDQ Academy (see Appendix A.4) is an example of a training program for developing knowledge and competencies for data.*

Table 6-10: Success criteria, recommended practices, justificatory knowledge, and evolution of the "People, Roles, and Responsibilities" design area

Success criteria	Recommended practices	Justificatory knowledge	Evolution
4a. Roles, responsibilities, and decision-making rights in connection with data are embedded in existing structures, defined, and actively managed.	− Define and update data-related roles, responsibilities, and decision-making rights − Integrate data-related roles and responsibilities into existing roles (e.g. in business process organization) − Define and implement mediation mechanisms to resolve conflicts between stakeholders − Define and ensure clear ownership for data	EDQM Maturity Model criterion 3a (Ofner, Otto et al., 2013, p. 14) IS organization management capability C 2.2 (Bärenfänger, 2017, p. 156) Data ownership as focus area of MDM (Spruit & Pietzka, 2015, p. 1072) Roles & responsibilities as a component in the Environmental Factors Hexagon (DAMA, 2017, p. 36) "A governance structure ensures consistent adherence to policies and procedures." "Data quality roles and responsibilities are documented for all staff involved in the data lifecycle." (PIC, 2016, p. 9). Organizational structures as an enabler domain of data governance (IBM Data Governance Council, 2007) Staff roles & skill set, data owners & stakeholders, governance organizational structure, roles & responsibilities, reporting alignment, and governance model & decision process as capabilities (GS1, 2010) Data governance as a Forrester data management capability (Hopkins et al., 2018) Information governance, organization and roles as enterprise information management building blocks (Gartner, 2014) Bosch case, Bayer 2 case, SBB case, tesa case	Extended

Success criteria	Recommended practices	Justificatory knowledge	Evolution
		DXM Maturity Assessment criterion 2.1 (see *Appendix C.1*)	

4b. A central data management organization coordinates and facilitates data-related activities across the organization.	− Establish and maintain a central data management organization to
	  o coordinate and facilitate activities of the federated data community
	  o develop, maintain, promote, and control corporate policies, standards, and guidelines for data management
	− Define interactions and establish reporting lines and managerial authority to coordinate roles in federated data organizations
		A control function designed to ensure custodial care of data as an enabler domain of data governance (IBM Data Governance Council, 2007)
		Data governance office as capability (GS1, 2010)
		Schaeffler case, Bosch case
			Definition

4c. Awareness of the importance of data is raised and continuously maintained across the entire organization.	− Communicate data strategy, data principles, and importance of excellent data − Make data citizens aware that they have duties and rights − Facilitate active participation of employees in data activities and eliminate resistance to necessary changes	EDQM Maturity Model criterion 3b (Ofner, Otto et al., 2013, p. 14) Data-driven culture as a key intangible resource (Gupta & George, 2016, p. 1051) Organization & culture as a component in the Environmental Factors Hexagon (DAMA, 2017, p. 36) Stakeholder engagement as a key aspect of a data management program (EDM Council, 2018) Awareness as an enabler of data governance (IBM Data Governance Council, 2007) Education & awareness, internal communication, and change management as data quality capabilities (GS1, 2010) Stakeholder engagement as a Forrester data management capability (Hopkins et al., 2018) tesa case, SAP 1 case	Adoption

Success criteria	Recommended practices	Justificatory knowledge	Evolution
		DXM Maturity Assessment criterion 2.4 (see *Appendix C.1*)	
4d. People are empowered to assume responsibility for data.	– Ensure individuals acquire the knowledge, skills, and information required to manage and use data – Run a training and education program to develop knowledge and competencies for data – Manage recruitment, career development, and succession planning of data-related roles – Identify and offer opportunities to talk about problems and share experiences and best practices regarding data (e.g. establish and manage a network of data experts).	EDQM Maturity Model criterion 3c (Ofner, Otto et al., 2013, p. 14) Big data analytics talent capability (Akter et al., 2016) Education and training of managerial and technical skills as a big data capability (Gupta & George, 2016, p. 1051) "Training is provided to ensure consistent adherence to procedures. Data quality standards are incorporated into individual performance targets to enhance accountability" (PIC, 2016, p. 9). Education as a key aspect of a data management program (EDM Council, 2018) Personal objectives, training, and job aids & work instructions as data quality capabilities (GS1, 2010) Data can only be managed effectively if data ownership and data stewardship are trained and actually executed (Khatri & Brown, 2010). SAP 1 case, Bosch case, CDQ Academy example DXM Maturity Assessment criterion 2.5 (see *Appendix C.1*)	Adoption

Activities in the People, Roles, and Responsibilities design area result in a data organization model, change management plan, information campaign, and skill development and education plan (see *Table 6-11*).

Table 6-11: Key result documents and supporting artifacts of the "People, Roles, and Responsibilities" design area

Key result documents	Supporting artifacts
– Data organization model (including role and board profiles, interaction models, RACI matrix) – Change management plan (for establishing data-driven culture) – Information campaign (e.g. website on the company's intranet, information brochure, newsletter, data management event plan – Skill development and education plan (including training materials)	Data governance organization layout (Otto & Reichert, 2010; Weber, 2009, p. 22; Weber, Otto, & Österle, 2009a), Reference model for data governance (Reichert, 2015; Weber, 2009, p. 106 et seqq.; Weber et al., 2009b), Data governance organization model (DAMA, 2017, p. 568) (Redman, 1996, 274ff; Yuhanna et al., 2011) Definition of data catalog roles (Korte et al., 2018), Overview of the Chief Data Officer role (Griffin, 2008; Horlacher & Hess, 2016), Overview of the Data Scientist and Solution Designer roles (Davenport & Patil, 2012)

6.5.5 Processes and Methods

Definition: Processes and methods define procedures and standards for proper and consistent data management.

Capabilities are implemented through organizational routines (Marino, 1996). As for data management, these routines include data management processes and methods. Data management processes define relevant data management procedures on a strategic, governance, and operational level and specify which tasks are to be executed by whom and in what order (Reichert, 2015, p. 58). *Strategic* data management processes define how a data strategy is to be developed and maintained. *Governance*-related data management processes define the standards for operational data management activities; in particular, these standards specify how data lifecycle activities are to be conducted, how data applications and the data architecture are to be managed, and how a performance management system is to be operated. *Operational* data management processes[40] define procedures for supporting activities, such as training, user and project support, and performance reporting (Reichert et al., 2013).

[40] In line with Design Decision 5 ("emphasize the data lifecycle") and contrary to the definitions of the Processes and Methods design area in the CDQM Framework (Otto et al., 2011), the EDQM Maturity Model (Ofner, Otto et al., 2013; EFQM, 2016), and Reichert's (2015) MDM process reference model, the DXM does not consider data lifecycle processes as part of data management processes. These are subject to a separate design area: Data Lifecycle.

Data management methods further specify and guide data-related activities, with the objective of ensuring standardized, enterprise-wide actions regarding data management and data use (Khatri & Brown, 2010). Examples of data management methods are corporate policies, standards, guidelines, standard operating procedures, or handbooks.

Excellence statement 5: Excellent organizations develop and maintain processes and methods for efficient data management that fully satisfy diverse stakeholders' expectations and requirements.

Two success criteria are implemented by organizations, which are excellent with regards to the Processes and Methods design area (see Table 6-12):

a. Excellent organization design, manage, and continuously improve data management processes on strategic, governance, and technical level and issue corporate policies, standards, guidelines, and methods for data management (e.g. governance and data structures, data models, data migration guidelines).

> *SBB's governance policy (see Subsection 5.4.3) specified relevant data management processes and methods.*

> *Bayer's process landscape (see Figure 5-7 on p. 106) defined data management processes, which are based on Reichert's (2015, p. 59) reference process model and were modified to reflect the capability view and the Goals-Enablers-Results structure of the DXM.*

> *Similarly, Bosch's data strategy (see Subsection 5.4.5) defined data management processes on strategic, governance, and operational level.*

b. Excellent organizations assure alignment between data management and business by embedding data management activities into business processes, corporate policies, and standards.

> *Bayer's corporate master data management directive (see Subsection 5.4.2) had references to the corporate data privacy and data protection directive of Bayer; et vice versa: the data privacy and data protection directive also included reference to the MDM directive and made adherence to the data management processes mandatory.*

> *SAP (see Subsection 5.3.3) integrated a data ethics survey into the marketing department's funneling process for data analytics scenarios.*

Table 6-12: Success criteria, recommended practices, justificatory knowledge, and evolution of the "Processes and Methods" design area

Success criteria	Recommended practices	Justificatory knowledge	Evolution
5a. Data management processes and methods are designed, managed, and continuously improved.	– Design and update data management processes on strategic, governance, and technical level – Develop and maintain corporate policies, standards, guidelines, and methods for data management (e.g. governance and data structures, data models, data migration guidelines)	EDQM Maturity Model criterion 4a (Ofner, Otto et al., 2013, p. 14) IS process management capability C 2.2 (Bärenfänger, 2017, p. 156). "Enterprise-wide policies and procedures established, documented, and communicated" (PIC, 2016, p. 8). Policies, standards, and operational procedures as a key aspect of a data governance (EDM Council, 2018) Policies as an enabler domain of data governance (IBM Data Governance Council, 2007) Documentation management, documentation standards, policies & standards management, operating procedures, process flow diagrams, and policy & standards review as data quality capabilities (GS1, 2010) SBB 1 case, Bayer 2 case, Bosch case DXM Maturity Assessment criteria 3.1 and 3.2 (see *Appendix C.1*)	Adoption
5b. Data management activities are embedded in business processes, corporate policies, and standards.	– Identify data users and relevant stakeholders, and work to meet their needs and expectations – Amend existing corporate policies, standards, and guidelines to integrate data-related activities in business routines	Activities and techniques in (business) processes as components in the Environmental Factors Hexagon (DAMA, 2017, p. 36) "Enterprise-wide policies and procedures adopted as	Definition

Success criteria	Recommended practices	Justificatory knowledge	Evolution
	− Design and maintain data provision and data management services to support data users (e.g., when performing mass data changes or data cleansing operations) and monitor the service level − Identify, continuously improve, and actively maintain the management of data in business processes	standard business procedures" (PIC, 2016, p. 9). (Business) processes involved in the information's lifecycle as a data quality capability (GS1, 2010) Bayer 2 case, SAP 2 case	

The Processes and Methods design area generates data management process documentations, corporate policies, standards, guidelines, and methods as key result documents (see *Table 6-13*).

Table 6-13: Key result documents and supporting artifacts of the "Processes and Methods" design area

Key result documents	Supporting artifacts
− Data management process documentations − Corporate policies, standards, guidelines, and methods for data management	Reference process model for master data management (Reichert, 2015, p. 59) MDM process management (Power, 2009) Methodology for data quality-oriented modeling and analysis of business processes (Ofner, Otto, & Österle, 2012)

6.5.6 Data Lifecycle

Definition: The data lifecycle manages all processes regarding the creation, acquisition, storage, maintenance, use, archiving, and deletion of data; defines and documents data (objects), data sources, data supply chains, data consumers, and data use contexts.

Business objects describe the core entities an enterprise needs to deal with in order to pursue its operations. Typical core business objects are suppliers, products, employees, and customers. The business objects are represented by data objects. For a common understanding, each business object and the data object it is represented by are clearly and unambiguously defined. Each data object has its own lifecycle, which outlines the stages "from cradle to grave". Managing the data lifecycle ensures a consistent, enterprise-wide approach to create, acquire, store, maintain, use, archive, and delete data (Redman, 1996; Wang, 1998; Wang et al., 1998). The data lifecycle process specifies the data lifecycle by detailing the relevant tasks and their order of appearance within each lifecycle stage, as well as the roles involved (Ofner, Straub, Otto, & Oesterle, 2013). Designing efficient lifecycle processes, which fulfill the demands of data consumers in multiple business functions, requires tracing data usage across the entire organization.

Excellence statement 6: Excellent organizations define and efficiently manage all relevant data objects, their respective lifecycles, and use.

Organizations being excellent in the Data Lifecycle design area demonstrate to three success criteria (see *Table 6-14*):

a. Excellent organizations unambiguously define, manage, and continuously improve key data objects and their lifecycles to comply with regulatory provisions, business rules, and data producers' and data users' requirements.

> *Both SBB's governance policy for data management (see Subsection 5.4.3) and Bayer's corporate master data management directive (see Subsection 5.4.2) included a chapter which defined the data lifecycle of core data objects and its stages.*

> *The Bosch data strategy (see Subsection 5.4.5) defined the stages of the data lifecycle process (i.e. onboarding, maintenance, synthesis, provisioning, archiving, and deletion) and provided principles for a methodology to establish and maintain data lifecycle processes.*

b. Excellent organizations trace where and how data is used across the organization to ensure data is handled in compliance with business rules and regulations and improve data lifecycles based on business requirements.

> *PMI's data catalog (see Subsection 5.3.1) comprises one module, which provides information on data lineage and data systems allowing to trace the data use in PMI's operations..*

> *To assign the data management roles from Bayer's corporate master data management directive (see Subsection 5.4.2), the master data management*

> *team reviewed data use in all IBO business processes on an attribute level. Based on this analysis, ownership for data objects/attributes was assigned to the involved business process owners.*

Table 6-14: Success criteria, recommended practices, justificatory knowledge, and evolution of the "Data Lifecycle" design area

Success criteria	Recommended practices	Justificatory knowledge	Evolution
6a. Data objects and their lifecycles are unambiguously defined, systematically documented, and actively managed.	– Ensure data object definitions are clearly and unambiguously documented and accessible across the entire organization – Document and model data lifecycles and principles (i.e. data creation, acquisition, storage, maintenance, usage, archiving, and deletion) for a better understanding of the use of data within the organization – Design, implement, monitor, and continuously improve data lifecycle to comply with operational excellence targets, regulatory provisions, business rules, and data producers' and data users' requirements	EDQM Maturity Model criterion 4c (Ofner, Otto et al., 2013, p. 14) Data processing / IS value chain capability B 1 (Bärenfänger, 2017, p. 156) Data lifecycle as focus area of MDM (Spruit & Pietzka, 2015, p. 1072) Information lifecycle management as a core discipline of data governance (IBM Data Governance Council, 2007) Initial data entry & setup and ongoing data maintenance as data quality capabilities (GS1, 2010) Data processing as a key DQM process (ISO, 2011) Business data services delivery as a Forrester data management capability (Hopkins et al., 2018) Documentation and understanding of the flow of information as an enterprise information management building block (Gartner, 2014) SBB 1 case, Bayer 2 case, Bosch case DXM Maturity Assessment criteria 4.1, 4.2, 4.3, and 5.4 (see *Appendix C.1*)	Modification
6b. Data use is traced, documented, and monitored.	– Identify all use cases and users of data (e.g. data consumption in business processes, data as input for report generation)	EDQM Maturity Model criterion 4b (Ofner, Otto et al., 2013, p. 14)	Modification

Success criteria	Recommended practices	Justificatory knowledge	Evolution
	− Document and monitor the use of data within the organization − Ensure data use cases comply with regulatory provisions and business rules	Data usage and data access as focus areas of MDM (Spruit & Pietzka, 2015, p. 1072) Model and manage information supply chains (Otto & Ofner, 2010). PMI case, Bayer 2 case	

Activities in the Data Lifecycle design area result in documentations of data (objects) and their respective lifecycle as well as documentations of data products, their respective data supply chains, and lifecycles (see *Table 6-15*).

Table 6-15: Key result documents and supporting artifacts of the "Data Lifecycle" design area

Key result documents	Supporting artifacts
− Data (object) glossary − Data lifecycle documentation (including principles) − Documentation of data traces/lineage	Method for master data integration: Description of business object types (Schmidt, 2010, p. 105 et seqq.), A Method for the identification and definition of information objects (Schmidt & Otto, 2008) Reference process model for master data management (Reichert, 2015, p. 59), Management of the master data lifecycle (Ofner, Straub et al., 2013), Model for information supply chain management (Otto & Ofner, 2010)

6.5.7 Data Applications

Definition: Data applications is about planning, implementing, and maintaining software which is designed to manage data and data products in order to achieve and maintain data excellence.

Achieving and maintaining data excellence is a task that cannot be encountered by simply implementing a suitable software solution. Therefore, data application management describes the planning, implementation, and maintenance of applications by which data management processes and data lifecycle processes are performed (Ballou et al., 1998). To manage data applications successfully, knowing the enterprise's application landscape and understanding which application performs which task and offers which function is key (Otto et al., 2012). A data application landscape documentation provides an overview of the applications in place, their interfaces, storage databases, and functions (Akter et al., 2016; Bourdreau & Couillard, 1999; Sun et al., 2006). Interfaces specify which other applications provide/receive which data to/from the applications documented. Furthermore, the type of interface (i.e. function-oriented or data-oriented interface), the requirements regarding response time, availability, and topicality, as well as integration patterns, frequency patterns, and initiation times are to be detailed. Storage databases define the data repository containing the data required for executing the applications' discrete functions (Österle, Höning, & Osl, 2011, p. 82 et seqq.). Functions document what tasks and processes the application is able to execute. Relevant functions in data management are data lifecycle management, data quality management, metadata management, and data integration among others (Otto et al., 2012). Core applications of data-driven enterprises include master data management (MDM) applications, data quality management (DQM) applications, and data catalogs.

Excellence statement 7: Excellent organizations provide and maintain the applications required to efficiently manage and use data.

Four success criteria are implemented by organizations, which are excellent with regards to the Data Applications design area (see *Table 6-16*).

a. Excellent organizations plan, actively managed, and continuously optimize their application landscape. They identify functions required and select adequate solutions from an ecosystem of various internal and external systems, applications, and/or tools. They document the (current and planned) application landscape as well as the functions and interfaces of each application.

> *For operationalizing its data strategy from a technological perspective, Bosch (see Subsection 5.4.5) reviewed the data management capabilities needed to implement the prioritize use cases. The capabilities were translated into data application functions required. These were aggregated in a data management application architecture. The architecture serves as the basis for selecting appropriate software solutions (already available or to be purchased).*

> *The corporate master data management directive of Bayer (see Subsection 5.4.2) outlined the core applications for conducting master data management. Furthermore, the responsibilities for managing the application landscape are defined: The Master Data Governance Sub-Council specified*

> *functional requirements, while Bayer's IT department selected appropriate applications and took care of their maintenance.*

b. The applications of excellent organizations provide an integrated and consistent version of core data objects as an enterprise-wide reference (i.e. "golden record"). Often, a SSOT is implemented through an MDM application as the central hub for managing the lifecycle of core data objects.

> *PMI (see Subsection 5.3.1) implemented one central master data hub which a number of systems have access to. The Data Management team selected SAP MDG, which was connected to two main target systems: PMI's ERP system and the commercial analytics application.*

c. Excellent organizations have data applications in place, which provide functions for reactive and proactive data quality management. They are able to prevent the creation of defective data (i.e. "first time right"), monitor data quality, identify and cleanse defective entries, and enrich datasets.

> *As the focus of this dissertation is on strategic data management, none of the ten case studies documented had a focus on quality-oriented data management in the Data Applications design area. However, several academic publications of the CC CDQ illustrate DQM applications in use: for example, at FMCG company Beiersdorf (Hüner, Schierning et al., 2011; Otto et al., 2012), automotive supplier Festo (Otto, 2012b; Otto et al., 2012), or a chemicals company (Otto, Ebner et al., 2010).*

d. Excellent organization create transparency regarding the data that is available within a company. They provide structural, governance, and content-related descriptions and allow access to this data through a data catalog application. As curated platforms, data catalogs match data supply (i.e. cataloged corporate data) and data demand (i.e. data discovery and access) (Korte et al., 2018).

> *The PMI Data Architecture team (see Subsection 5.3.1) maintained Collibra as the company's platform creating transparency regarding all data available. The data catalog application provided data definitions, data structures, and data sources, as well as data-related rules and responsibilities to support data governance. It was constantly updated and populated every time a business function was in contact with EAD. Collibra comprised two main modules: (1) the business glossary, providing high-level information to business users who were looking for specific information such as definitions of a*

> *certain term, a certain data entity/attribute, or a certain KPI; and (2) the data dictionary, providing technical information, such as data profiles or data lineage information, but also information on data systems and data models. Both modules were connected to each other, so that there was a bridge from business to IT allowing users to navigate from a business term directly to the corresponding data entity.*

Table 6-16: Success criteria, recommended practices, justificatory knowledge, and evolution of the "Data Applications" design area

Success criteria	Recommended practices	Justificatory knowledge	Evolution
7a. Application landscape is planned, actively managed, and continuously optimized.	− Identify requirements and activities that need to be supported by data organization − Test, evaluate, and select software applications from vendor base − Understand the application landscape for data management and data use as an ecosystem comprising various internal and external systems, applications, and/or tools − Document application landscape, including functions and interfaces of each application − Document and understand the gap between as-is and to-be application landscape, and manage a roll-out plan and roadmap for closing the gap between as-is and to-be	EDQM Maturity Model criteria 6a-c (Ofner, Otto et al., 2013, p. 14) Application management capability C 2.3 (Bärenfänger, 2017, p. 156) Tools as component in the Environmental Factors Hexagon (DAMA, 2017, p. 36) Design of physical architecture (incl. platforms and tools) as a key aspect of technology architecture (EDM Council, 2018) System design & architecture and system requirements documentation as data quality capabilities (GS1, 2010) Data management technology research and planning, data management technology implementation and maintenance, and data management technology operation as Forrester data management capabilities (Hopkins et al., 2018) Information infrastructure as an enterprise information management building block (Gartner, 2014) Bosch case, Bayer 2 case DXM Maturity Assessment criteria 5.2 and 5.3 (see *Appendix C.1*)	Modification

Success criteria	Recommended practices	Justificatory knowledge	Evolution
7b. Applications provide an integrated and consistent version of core data objects as an enterprise-wide reference (“golden record”).	– Provide an integrated and consistent version of core data objects as an enterprise-wide reference – Define the leading application for storage and for data lifecycle processes for each core data object and its attributes – Enrich core data objects based on data available in all (internal and external) sources	Managing the lifecycle is a core function of data applications (Otto et al., 2012). Defensive data management approaches require a single source of truth (DalleMulle & Davenport, 2017) PMI case DXM Maturity Assessment criteria 5.1 and 5.4 (see *Appendix C.1*)	Definition
7c. Applications provide functions to analyze, monitor, and continuously improve data quality.	– Analyze and profile data and its quality – Monitor data quality – Identify and cleanse defective data – Enrich data (e.g. link internal and external data to enable automatic validation, enhancement, and correction of data)	“Data quality issues are recognized early in the information flow. Technology enables early detection and correction” (PIC, 2016, p. 9). Data validation, workflow & user interface, security & access control, and revision & change history as data quality capabilities (GS1, 2010) Data quality management is a core function of data applications (Otto et al., 2012). Beiersdorf case (Hüner, Schierning et al., 2011; Otto et al., 2012), Festo case (Otto, 2012b; Otto et al., 2012), and Chemicals company case (Otto, Ebner et al., 2010) DXM Maturity Assessment criterion 6.4 (see *Appendix C.1*)	Definition
7d. Applications provide functions to describe, catalog, provide, and access data and their metadata for consistent usage across the organization.	– Make data FAIR, i.e. findable, accessible, interoperable, and reusable for people and machines – Define users and access rights, protect restricted data from unauthorized access considering data sovereignty	Unified data repository, external publication, and internal publication as data capabilities (GS1, 2010) Metadata management is a core function of data applications; metadata specifies data properties, data structures, and the meaning of data (Otto et al., 2012). FAIR data supports data discovery and use (Wilkinson et al., 2016).	Definition

Success criteria	Recommended practices	Justificatory knowledge	Evolution
		Data catalogs create transparency about data supply and demand (Korte et al., 2018). PMI case DXM Maturity Assessment criteria 6.2 and 6.5 (see *Appendix C.1*)	

Activities in the Data Applications design area result in an application landscape documentation, application descriptions, and applications for MDM, DQM, and data cataloging (see *Table 6-17*).

Table 6-17: Key results and supporting artifacts of the "Data Applications" design area

Key results	Supporting artifacts
− Application landscape documentation − Application descriptions (including functions and interfaces) − Applications for managing (master) data, managing data quality, and cataloging/curating data	Method for corporate data architecture design (Ebner, 2014, p. 103 et seqq.), Decision model for master data application architecture (Baghi et al., 2014), Application and integration architecture design (Österle et al., 2011, p. 82 et seqq.), Method for master data integration: application analysis (Schmidt, 2010, p. 147 et seq.) Functional reference model for CDQM (Otto et al., 2012), Data catalog functional reference model (Korte et al., 2018)

6.5.8 Data Architecture

Definition: The data architecture defines and maintains specifications that provide a shared business vocabulary, express strategic data requirements, and outline high-level, integrated architecture landscape designs and data flows.

The data architecture complements and connects the process view and the application view of data management. It provides a shared business vocabulary and outlines high-level, integrated application landscape designs and data flows (Goodhue et al., 1988). As a basis for data architecture, data management-related business rules are directives that govern, guide or influence the habits and procedures of business departments with regard to data management. They are characterized by suitability (i.e. the rules are written in a language business understands), actionability (i.e. readers directly understand what they are requested to do), and business ownership (i.e. the rules are created and maintained by business itself, as they know best about their requirements) (OMG, 2017a).

Data models depict the core business objects and their relations (Brancheau, Schuster, & March, 1989). They either depict data requirements from a business perspective in the form of conceptual data models (i.e. modeling the reality) or describe the design and management of data from an application perspective in the form of physical data models (i.e. modeling the technical representation) (Frank, 1994, 87 et seqq.). A *conceptual data model* provides an abstract, formal description of business objects and their relations. It is documented by means of a modeling language, such as the entity relationship model (Chen, 1976), and is often accompanied by a business (data) glossary, which documents metadata (such as definitions the objects and their attributes or synonyms) (Schmidt, 2010, p. 126). Having a corporate-wide conceptual data model is regarded as a key enabler of data management, as it is “a prerequisite for a uniform understanding of the data and therewith for the intended use of the data as well” (Otto & Österle, 2015, 27). The *physical data model* describes the data objects from an application view and documents technical metadata, such as data type, field length, or primary or foreign keys (Hüner, Otto, Österle, & Brauer, 2011, p. 99). As a third type of data models, a *logical data model* defines the structure and relationships between data objects. Thereby, it provides the basis for translating the conceptual data model into a concrete implementation of a physical data model.

As data is typically managed by more than one application, consistency of data across different applications is crucial. Consequently, both the leading applications for data storage and distribution and the consuming applications processing data objects and their attributes need to be documented (Schmidt, 2010, p. 160). Furthermore, as a lot of data is available either internally or can be obtained from external sources, the data sources which are used to acquire, validate, or enrich data objects need to be specified. By modeling data flows, the relationships between applications can be specified. Based on the static application landscape documentation and data object descriptions, data flow diagrams provide a dynamic view of the flow of data object attributes between applications (Schmidt, 2010, p. 169).

Excellence statement 8: Excellent organizations define core data objects through data models, document business rules, and design the data storage and distribution architecture for data being consistently understood and used across the entire organization.

Organizations being excellent in the Data Architecture design area apply three success criteria (see *Table 6-18*):

a. Excellent organizations develop, permanently assess, and communicate a conceptual data model to establish a common understanding of core business entities.

> *SBB's data strategy 2017-2020 (see Subsection 5.4.3) defined eleven core data domains. The leading data manager of each domain was required to design and update a conceptual data model for his/her domain.*

> *PMI's Data Architecture team (see Subsection 5.3.1) documented over 690 data entities and over 4,840 data attributes in Collibra. The team had developed and maintained 27 conceptual and 54 logical data models for several domains such as product, device, customer, consumer, and business partner.*

> *In Bayer's corporate MDM directive (see Subsection 5.4.2), key data objects are documented in a conceptual data model.*

b. Data storage, distribution, and flow is designed, implemented, and actively managed by excellent organizations for consistent data use across the entire organization.

> *Bayer's corporate MDM directive (see Subsection 5.4.2) specified which data object is stored in which application, and defined principles for how data flows between applications.*

c. Excellent organizations document business rules for consistent data handling across the entire organization.

> *PMI's data catalog (see Subsection 5.3.1) comprises one module, which provides information on data-related rules. It is constantly updated and populated every time a business function is in contact with EAD.*

Table 6-18: Success criteria, recommended practices, justificatory knowledge, and evolution of the "Data Architecture" design area

Success criteria	Recommended practices	Justificatory knowledge	Evolution
8a. Common understanding of a data model for the business entities is developed, permanently assessed, and communicated.	− Formalize, define, document, and manage business vocabularies in a standardized form to make them actionable − Identify core business objects embedded within the data strategy's scope and ensure agreement on unambiguous definitions of business objects and their metadata among key stakeholders − Formalize a conceptual (business) object model and its metadata based on common standards − Develop, maintain and publish a business data model to establish a common understanding of core business entities	EDQM Maturity Model criterion 5a (Ofner, Otto et al., 2013, p. 14) Data architecture capability C 2.1 (Bärenfänger, 2017, p. 156) Data model as a key aspect of MDM (Spruit & Pietzka, 2015, p. 1072) Data modeling & design and metadata management as knowledge areas (DAMA, 2017, p. 123 et seqq., p. 417 et seqq.) Data architecture as a key component (EDM Council, 2018) Data architecture and classification & metadata as supporting disciplines of data governance (IBM Data Governance Council, 2007) Data definitions & standards as a data quality capability (GS1, 2010) Data architecture management and data design as key DQM processes (ISO, 2011) Data architecture development as a Forrester data management capability (Hopkins et al., 2018) Core data entities and their relationships are described by data models (Brancheau et al., 1989) SBB 1 case, PMI case, Bayer 2 case DXM Maturity Assessment criterion 6.1 (see *Appendix C.1*)	Extension
8b. Data storage, distribution, and flow are designed,	− Define and document databases for storage of each data object and data attribute	EDQM Maturity Model criterion 5b (Ofner, Otto et al., 2013, p. 14)	Adoption

Success criteria	Recommended practices	Justificatory knowledge	Evolution
implemented, and actively managed.	− Define (both mandatory and optional) principles and guidelines for data-related activities to model, assess, and monitor data flow between applications − Define and document data flow across application landscape	Data storage as focus area of MDM (Spruit & Pietzka, 2015, p. 1072) Data storage & operations and data integration & interoperability as knowledge areas (DAMA, 2017, p. 169 et seqq., p. 269 et seqq.) Data acquisition, storage, integration, and distribution as key aspects of technology architecture (EDM Council, 2018) Data flow management as a key DQM process (ISO, 2011) Document leading applications for data storage and distribution as well as data consuming applications and their interfaces (Goodhue et al., 1988) Data flows are streamlined by documenting and monitoring interfaces and storage (Akter et al., 2016; Bourdreau & Couillard, 1999; Sun, Zhao, Nunamaker, & Sheng, 2006) Bayer 2 case DXM Maturity Assessment criterion 6.3 (see *Appendix C.1*)	
8c. Business rules are systematically documented and managed.	− Review business processes for data management related requirements − Formalize, define, document, and manage business rules in a standardized form to make them actionable	Manage business rules to ensure data quality (Schlosser, 2017, p. 141 et seqq.; Schlosser, Baghi, Otto, & Oesterle, 2014). Data security & use policy as a data quality capability (GS1, 2010) PMI case	Definition

The Data Architecture design area provides a business vocabulary and rulebook, conceptual (business) data models, a data storage and distribution architecture, and data flow diagrams (see *Table 6-19*).

Table 6-19: Key result documents and supporting artifacts of the "Data Architecture" design area

Key result documents	Supporting artifacts
− Business vocabulary and rulebook − Conceptual, logical, and physical data models − Data storage and distribution architecture − Data flow diagrams	Business rules management (Schlosser et al., 2014), Implementation of the business vocabulary and rulebook (Schlosser, 2017, p. 141 et seqq.), Meta data management in a semantic wiki (Hüner, Otto, Österle et al., 2011) A method for the identification and definition of information objects (Schmidt & Otto, 2008), Conceptual/semantic data model development (Schmidt, 2010, p. 140 et seqq.), Method for corporate data architecture design (Ebner, 2014, 103 et seqq.), Data modeling and design (DAMA, 2017, p. 123 et seqq.) Data storage design (DAMA, 2017, p. 169 et seqq.) Strategic data planning method (Goodhue et al., 1988)

6.5.9 Performance Management

Definition: Performance management plans, implements, and controls all activities for measuring, assessing, improving, and ensuring data management performance, data excellence, and business value.

Managing data as a strategic resource requires activities to measure, assess, improve, and ensure the progress and outcome of managing this resource (Ferreira & Otley, 2009). Performance management orchestrates these activities (Möller, Wirnsperger, & Gackstatter, 2015) in order to answer five questions (Otley, 1999, p. 365 et seq.):

1. How can the achievement of key objectives be evaluated?
2. How can the performance of activities for achieving objectives be assessed and measured?
3. What are the desired target values with regard to objectives and activities?
4. What kind of reward should employees be granted for meeting the performance targets set, and what kind of penalty should be imposed for missing targets?
5. How can learning from experience be enabled?

While question no. 4 is addressed by the People, Roles, and Responsibilities design area of the DXM (see *Subsection 6.5.4*), and question no. 5 by Continuous Improvement (see *Subsection 6.5.12*), the first three questions need to be answered by the Performance Management design area of the DXM. In considering data as a strategic resource that has high business relevance, two general goals of data management are of importance: achieving data excellence (see *Subsection 6.5.10* for details) and generating business value (see *Subsection 6.5.11* for details).

Developing company-specific metrics for data excellence and business value is the initial step for answering the first question. Defining process-oriented data management metrics (Ofner, 2013) and metrics for evaluating data management maturity and progress (Ofner, Otto et al., 2013) addresses question no. 2. Question no. 3 can be answered by specifying target values for each metric introduced before. Finally, the performance management system specifies the processes for defining, measuring, controlling, aggregating, and reporting the metrics (Baghi, 2016).

Excellence statement 9: Excellent organizations define measures to continuously review and improve data excellence, performance, progress, and business value of all data management activities.

Three success criteria are addressed by organizations, which are excellent with regards to the Performance Management design area (see Table 6-20).

a. Excellent organizations review the impact of data, data use cases, and data management on business performance. They define and actively management business value metrics.

> *PMI (see Subsection 5.3.1) defined business value metrics (i.e. financial value generated through the EAD organization and, especially, its use cases delivered) and target values (i.e. generate five times more business value than EAD cost).*

> *Schaeffler's Corporate Data Management (see Subsection 5.4.4) defined measures to track the impact of its data management activities on business processes and the time/costs saved through a reduction of service processing time.*

b. Based on business value objectives, excellent organizations define and actively manage data excellence metrics.

> *Schaeffler (see Subsection 5.4.4) maintained a comprehensive KPI methodology to continuously monitor and improve data quality per domain.*

> *SAP (see Subsection 5.3.3) developed a qualitative assessment method for analyzing whether new big data scenarios comply with SAP's ethical standards.*

> *The Bosch data strategy (see Subsection 5.4.5) defined data excellence as core objective: "We establish data rules and policies that provide the best data quality, security, and control for our customers to establish RB as the most trustful IoT partner." Consequently, the strategy implementation roadmap includes the development of data quality, data security, and data compliance/control metrics.*

c. Excellent organizations define and actively manage performance and progress metrics for their data management activities.

> *Schaeffler's Corporate Data Management (see Subsection 5.4.4) defined (among others) the cycle times of its data management services to business stakeholders as a KPI.*

> *PMI's Data Management team (see Subsection 5.3.1) developed its own maturity assessment model to evaluate data maturity in all of PMI's business functions, thereby creating transparency with regard to the status quo and providing a baseline for tracking EAD's progress.*

> *The maturity assessment questionnaire (see Appendix C.1) developed by the author of this dissertation and SBB (see Subsection 5.5.1) provides an example of what data management maturity metrics may look like. SBB had applied the revised maturity model for eight data domains already (with more maturity assessments planned for the future).*

> *Finally, Schaeffler (see Subsection 5.4.4) participated in (and won) the CDQ Good Practice Award in 2016. The award submission can be regarded as a means of externally assessing a data management approach's maturity.*

Table 6-20: Success criteria, recommended practices, justificatory knowledge, and evolution of the "Performance Management" design area

Success criteria	Recommended practices	Justificatory knowledge	Evolution
9a. Business value metrics are defined, actively managed, and measured.	− Specify business value metrics (e.g. scales, points of measurement, methods of measurement) based on cause-and-effect relationships between data defects/ improvements/usage and business performance indicators − Track data use cases and the business value they generate (e.g. additional revenue, efficiency gains)	EDQM Maturity Model criteria 2a and 10a (Ofner, Otto et al., 2013, p. 14) Data quality impact on business as a focus area of MDM (Spruit & Pietzka, 2015, p. 1072) Monitor impact of erroneous data as a data quality capability (GS1, 2010) PMI case, Schaeffler case DXM Maturity Assessment criterion 7.1 (see *Appendix C.1*)	Modification
9b. Data excellence metrics are defined, actively managed, and measured.	− Identify and define data excellence dimensions regarding data quality, data privacy, data compliance, and/or data security − Specify data excellence metrics (e.g. scales, points of measurement) and reasonable threshold and target values − Develop, implement, and continuously improve methods for measuring data excellence	EDQM Maturity Model criteria 2a and 2c (Ofner, Otto et al., 2013, p. 14) DQM capability C 2.1 (Bärenfänger, 2017, p. 156) Data quality assessment and data protection as focus areas of MDM (Spruit & Pietzka, 2015, p. 1072) Data security and data quality as knowledge areas (DAMA, 2017, p. 217 et seqq., p. 449 et seqq.) Data quality as well as privacy and security as key components and aspects (EDM Council, 2018) Data quality management and information security & privacy as core disciplines of data governance (IBM Data Governance Council, 2007) Workflow controls, system validation, audit procedures, risk management, and customer feedback policy as data quality capabilities (GS1, 2010) Data quality planning, data quality criteria setup, and data quality measurement as	Modification

Success criteria	Recommended practices	Justificatory knowledge	Evolution
		key DQM processes (ISO, 2011) Data security as a Forrester data management capability (Hopkins et al., 2018) Schaeffler case, SAP 2 case DXM Maturity Assessment criteria 7.2 and 7.3 (see *Appendix C.1*)	
9c. Data management performance and progress metrics are defined, actively managed, and measured.	– Specify performance and progress metrics for data management taking expectations and requirements of different stakeholders into account – Define reasonable threshold and target values for performance and progress of data organization – Develop, implement, and continuously improve methods of measurement of performance and progress metrics	EDQM Maturity Model criterion 10b (Ofner, Otto et al., 2013, p. 14) Audit information, logging & reporting as a supporting discipline of data governance (IBM Data Governance Council, 2007) Performance management and performance metrics (GS1, 2010) PMI case, SBB case, Schaeffler case DXM Maturity Assessment criterion 7.4 (see *Appendix C.1*)	Modification

Activities in the Performance Management design area result in a performance management system for data management and in metrics for business value, data excellence, data management performance, and data management progress (see *Table 6-21*).

Table 6-21: Key result documents and supporting artifacts of the "Performance Management" design area

Key result documents	Supporting artifacts
– Performance management system – Metrics for business value, data excellence, performance and progress	Usage-based data valuation (Möller, Otto, & Zechmann, 2017), Balanced score card (Kaplan & Norton, 1996), Benefits dependency network (Ward & Daniel, 2006, p. 133 et seqq.) Capability reference model for data quality controlling (Baghi, 2016), Data quality controlling systems (Hüner, 2011), Measuring data quality (Otto, Ebner et al., 2010), Data quality management metrics (DAMA, 2017, p. 449 et seqq.)

Key result documents	Supporting artifacts
	Data protection and data access metrics (Spruit & Pietzka, 2015), Data security metrics (DAMA, 2017, p. 217 et seqq.)

Legner et al. (2017, p. 40 et seqq.) provide examples for business value, data excellence, data management performance and progress metrics. Based on their overview, *Table 6-22* lists relevant metrics, short descriptions, and typical measurement and reporting methods.

Table 6-22: Short description, typical measurement and reporting methods of data management metrics

Metric type	Metric	Description	Typical measurement/reporting method
Business value metrics	Impact on strategic goals	Impact of data management on strategic business goals	Assessed qualitatively and visualized by means of dependency graphs (Peppard, Ward, & Daniel, 2007) or traffic light charts
	Economic value of data	Financial value of data	Assessed by means of the reproduction cost approach or the use-based approach (Möller et al., 2017; Zechmann, 2016)
	Impact on business process related goals	Impact of data management on business process KPIs	Visualized by means of dependency graphs or traffic light charts
	Cost/time savings	Cost/time savings due to more efficient data maintenance processes, automated data cleansing/data import processes	Assessed by means of process mining
	Satisfaction of external groups	Satisfaction of customers, consumers, or business partners with respect to data excellence (e.g. quality of product catalogs, quality of shared data, adherence to data privacy standards and consents)	Surveyed by means of questionnaires/ interviews
Data excellence metrics	Data quality	Quantitative assessment of data's "fitness for use" (e.g. consistency, completeness, or accuracy)	Measured in terms of conformance of data with respect to certain data quality dimensions

Metric type	Metric	Description	Typical measurement/reporting method
	DQ Audit findings	Number of corporate data quality related violations during an audit (e.g. ISO 9001:2008)	Measured by reviewing audit results
Data management performance metrics	Cycle/ turnaround time	Time passed from requesting a new master data object (i.e. a new supplier or consumer data record) until this record is available in operational systems (e.g. ERP)	Measured by process mining, workflow logs, or ticketing system logs
	Internal satisfaction	Satisfaction of company-internal stakeholders such as data requestors and consumers in business processes	Surveyed by means of questionnaires/interviews
Data management progress metrics	Maturity score	Maturity assessment of current capabilities from a strategic, organizational and technical point of view	Surveyed by means of questionnaires/interviews
	Supported use cases	Percentage of agreed use cases fully supported by data management	Tracked by means of a use case funnel
	Rulebooks	Percentage of data domains covered by rulebooks (i.e. definitions, data models, processes, roles, responsibilities, methodologies).	Measured by means of a gap analysis between rulebook and data model
	Data records under governance	Percentage of data records covered by detailed rules	Measured by means of a gap analysis between rulebook and data model
	Geographical regions/ branches	Percentage of geographical regions/ branches implementing data governance	Measured by means of achieved milestones in rollout plans
	Role assignments	Percentage of geographical regions/branches implementing data governance.	Measured by means of achieved milestones in rollout plans
	Trained people	Percentage of roles assumed by appropriately trained people	Measured by means of achieved milestones in rollout plans

6.5.10 Data Excellence

Definition: Data excellence refers to the impact of data management on data quality (defined as "fitness for purpose"), but also on additional data-related dimensions, such as regulatory compliance, data security, or data privacy.

Being an outcome-oriented capability, data management has a direct impact on data itself, which the DXM defines as "Data Excellence". The concept of excellence originated from approaches of TQM (Suarez et al., 2016) and was transferred to the data domain by EFQM (EFQM, 2016). Like quality-oriented data management, data excellence comprises the goal of providing high-quality data (Batini & Scannapieca, 2006; English, 2003; Wang, 1998; Wang et al., 1998), but it also addresses additional data-related dimensions, such as data privacy, data security, or compliance (Delbaere & Ferreira, 2007; Sadeghi et al., 2015). What dimensions are actually of relevance in the context of data excellence is up to each enterprise to decide.

- Data quality is a multi-faceted concept comprising various aspects, such as consistency, completeness, accuracy, or timeliness (Wang & Strong, 1996). As these aspects are context-specific and depend on the perception and expectation of the data user, data quality is defined as data's "fitness for purpose" (Otto, 2011b; Redman, 2001; Wang & Strong, 1996).

- Data privacy describes the challenge of handling personal and other sensitive data, while considering general data protection laws and individuals' data privacy preferences (Torra, 2017).

- Data security aims at protecting data from unauthorized access to meet regulations, contractual agreements, or business requirements (DAMA, 2017, p. 217).

- Data compliance aims at meeting the requirements of data-related laws or regulations (such as GDPR) as well as the expectations of data providers/owners in the sense of data sovereignty and data ethics (DAMA, 2017, p. 49 et seqq.; Delbaere & Ferreira, 2007).

As transparency regarding the outcome of data management is the basis for continuous improvement (Batini et al., 2009; EFQM, 2009), it is important that the results and achievements regarding each relevant data excellence dimension are communicated across the entire organization. Furthermore, the overall performance and progress of data management (see previous *Subsection 6.5.10*) are to be reported as well.

Excellence statement 10: Excellent organizations comprehensively report on data excellence achievements as well as on performance and progress of all data management activities.

Organizations being excellent in the Data Excellence design area demonstrate two success criteria (see *Table 6-18*):

a. Excellent organizations appropriately communicate data excellence achievements to all relevant stakeholders.

> *Company-specific data excellence metrics are already presented in the previous Subsection 6.5.10. The companies mentioned (i.e. Schaeffler and SAP) do not only define but also measure and communicate the data excellence achievements. For example, Schaeffler reported a data quality of 98.8 percent within the "debtor" data domain.*

b. Excellent organizations appropriately communicate performance and progress of data management to all relevant stakeholders.

> *Schaeffler's Corporate Data Management (see Subsection 5.4.4) measured and communicated the cycle times of its data management services to business stakeholders. Over time, the company observed a reduction of service processing time by 60 per cent.*

> *PMI's Data Management team (see Subsection 5.3.1) applies its maturity assessment on a constant basis: each data domain is assessed every six months.*

> *SBB (see Subsection 5.5.1) reviews the progress of its eleven core data domains on a regular basis. Every two years, a data domain's maturity is evaluated with the maturity assessment questionnaire (see Appendix C.1). The assessment results are compared to previous assessments and communicated to the executive board.*

Table 6-23: Success criteria, recommended practices, justificatory knowledge, and evolution of the "Data Excellence" design area

Success criteria	Recommended practices	Justificatory knowledge	Evolution
10a. Data excellence is continuously reported and appropriately communicated to all stakeholders.	– Report data excellence (i.e. data quality, data privacy, data compliance, data security) on a permanent basis – Identify relevant stakeholders and define the communication (channels) for each group – Create transparency regarding data excellence results of data management e.g. via reports, dashboards, or alerts	EDQM Maturity Model criterion 2b (Ofner, Otto et al., 2013, p. 14) Awareness of data quality gaps as a focus area of MDM (Spruit & Pietzka, 2015, p. 1072) Data risk management & compliance as an outcome domain of data governance (IBM Data Governance Council, 2007) Performance reporting on data quality and process compliance audits as data quality capabilities (GS1, 2010) Schaeffler case, SAP 2 case DXM Maturity Assessment criteria 8.1 and 8.2 (see *Appendix C.1*)	Extension
10b. Performance and progress of data management is continuously reported and appropriately communicated to all stakeholders.	– Report performance and progress of data management on a permanent basis – Identify relevant stakeholders and define communication (channels) for each group – Create transparency regarding performance and progress of data management, e.g. visualize the results of data management services and data products via dashboards	EDQM Maturity Model criterion 10b (Ofner, Otto et al., 2013, p. 14) Performance reporting on service levels, external & internal feedback, product measurements, and review & reporting of audit results as data quality capabilities (GS1, 2010) Schaeffler case, PMI case, SBB 2 case	Modification

The Data Excellence design area results in periodic reports on data excellence, data management performance, and data management progress (see Table 6-24).

Table 6-24: Key result documents and supporting artifacts of the "Data Excellence" design area

Key result documents	Supporting artifacts
− Periodic reports on data excellence, data management performance and progress	Capability reference model for data quality controlling (Baghi, 2016), Data quality controlling systems (Hüner, 2011)

6.5.11 Business Value

Definition: Business value refers to data's and data management's impact on business.

Through providing excellent data, data products, and data management services, data management creates value and (positively) impacts business performance (Joshi & Rai, 2000; Sheng & Mykytyn, 2002; Spruit & Pietzka, 2015; Xu et al., 2016). Data is monetized in terms of new or enhanced business models, operational excellence, data-driven insights, and decision-making (Wixom & Ross, 2017). Consequently, the business value of data materializes in (1) increased revenues and (2) reduced costs:

(1) Data-driven companies increase their revenues through expanding their value proposition with data-enriched products and services and data-driven services (Schüritz et al., 2017) and maintaining/expanding the number of satisfied customers through excellent data provided.

(2) Data-driven companies reduce costs through data-enabled process improvements (Schüritz et al., 2017), which aim to optimize processes and increase productivity, support decision-making (Orr, 199 (Davenport, Harris, Long, & Jacobson, 2001)8; Price & Shanks, 2005; Shankaranarayanan et al., 2003), and gain insights about customers (Kiron & Shockley, 2011).

Creating transparency regarding top-line and bottom-line impact of data and data management and communicating its value to business, helps to raise awareness for data throughout the enterprise (Chen et al., 2012; LaValle et al., 2011). Besides data's impact on a company's profit-and-loss, data itself (like any other tangible or intangible resource) has an inherent monetary value, which should be assessed and reported (Atkinson & McGaughey, 2006; Borek, Parlikad, Woodall, & Tomasella, 2014; Even & Shankaranarayanan, 2007; Moody & Walsh, 1999). Determining the financial value of data follows established accounting methods like the market-based, cost-based, or income-based valuation approaches (Möller et al., 2017; Zechmann, 2016).

Excellence statement 11: Excellent organizations comprehensively report on and achieve outstanding business value generated by data.

One key success criterion is demonstrated by organizations, which are excellent with regards to the Business Value design area (see *Table 6-25*): They continuously measure and appropriately communicate the business value generated by data (in general and per data use case) in terms of new or enhanced business models, operational excellence, data-driven insights and decision making, reduced risks, and ensured compliance.

> *PMI reported the business value of its data science use cases on a dashboard available to the entire EAD organization. The seven data use cases in preparation, the eight use cases in the delivery phase, and the one already industrialized represented a high double-digit number, in million USD, of potential costs savings or additional revenues.*

> *Similarly, Bosch (see Subsection 5.4.5) calculated the (expected) business value of the prioritized data use cases resulting in a significant financial value potential.*

> *Schaeffler's Corporate Data Management (see Subsection 5.4.4) tracked the impact of its data management activities on business processes and the costs saved in business through a reduction of the data management service processing time.*

Table 6-25: Success criterion, recommended practices, justificatory knowledge, and evolution of the "Business Value" design area

Success criteria	Recommended practices	Justificatory knowledge	Evolution
11. Business value generated by data and data management is continuously reported and appropriately communicated to all stakeholders.	− Report the business value generated by data and data management continuously − Identify relevant stakeholders and define communication (channels) for each group − Create transparency about the business value of data	EDQM Maturity Model criterion 10a (Ofner, Otto et al., 2013, p. 14) Data management business case as a key component (EDM Council, 2018) Value creation as an outcome domain of data governance (IBM Data Governance Council, 2007)	Modification

Success criteria	Recommended practices	Justificatory knowledge	Evolution
	generated (in general and per data use case) in terms of o operational excellence, o new or enhanced business models, and o risk reduction and enhanced compliance	Business case (metrics) as an enterprise information management building block (Gartner, 2014) PMI case, Bosch case, Schaeffler case DXM Maturity Assessment criteria 8.3 and 8.4 (see *Appendix C.1*)	

Activities in the Business Value design area generate periodic business value reports, which either communicate the business contribution of each data use case or data's business value in an aggregated form across all data management activities and data use cases (see *Table 6-26*).

Table 6-26: Key result documents and supporting artifacts of the "Business Value" design area

Key result documents	Supporting artifacts
− Periodic reports on business value (on aggregated level and per data use case)	Benefits dependency network (Ward & Daniel, 2006, p. 133 et seqq.) Financial data valuation (Möller et al., 2017; Zechmann, 2016)

6.5.12 Continuous Improvement

Definition: Continuous improvement allows adjusting the Goals and Enablers based on the Results achieved, ensuring the dynamic nature of data management.

Given the understanding of data as a strategic resource, and adopting the idea of the continuous management cycle, data management is understood as a continuous effort requiring constant adjustment. According to Wu & Chen, 2006), continuous improvement is characterized by four criteria: (1) an enterprise-wide focus on process performance improvements, (2) gradual improvements through step-wise innovation, (3) organization-wide activities involving all employees, and (4) the creation of a learning and growing environment. To implement continuous improvement, it is important to define routines for initiating improvements as well as techniques to document and control measures for continuous improvement (Bessant & Caffyn, 1997, p. 7 et seqq.).

Consequently, a continuous improvement process in the data management domain includes learning from Results achieved in order to adapt Goals and Enablers accordingly. Furthermore, it is important that all measures for continuous improvement are thoroughly documented and monitored.

> **Excellence statement 12:** Excellent organizations regularly review the results of data management and adjust their enablers to further increase the level of data excellence and the business value generated.

Organizations being excellent in the Continuous Improvement design area implement one key success criterion (see *Table 6-27*): They have a continuous improvement cycle for all data-related activities is in place, which initiates projects and/or trigger activities for adjusting the Goals or Enablers based on monitored Results. Excellent organizations offer opportunities to share experiences and good practices regarding data management within the organization and encourage individuals and teams to initiate and participate in improvement activities.

> *PMI (see Subsection 5.3.1) applied a maturity assessment for continuously improving its data management activities. Furthermore, the PMO continuous reviews the priorities of the EAD organization and adapts the EAD roadmap in case of changing external or internal conditions.*

> *Schaeffler (see Subsection 5.4.4) extended the scope of its Corporate Data Management by transferring established concepts to further data domains, which were regarded as relevant for future activities.*

> *SBB's maturity assessment activities (see Subsection 5.5.1) aimed at improving a data domain's maturity within a period of two years. After a presentation of the maturity assessment results of a certain domain, the leading data domain manager had to define target values to be met in the next assessment in two years.*

Table 6-27: Success criterion, recommended practices, justificatory knowledge, and evolution of the "Continuous Improvement" design area

Success criteria	Recommended practices	Justificatory knowledge	Evolution
12. Continuous improvement cycle encompassing all data management enablers is in place.	– Define a process for continuous improvement (i.e. involved roles, responsibilities, sequence of activities, escalation path) – Initiate projects and/or trigger activities for improvement based on performance monitoring and threshold violations – Define and review priorities of improvement projects and activities based on data excellence and business value considerations – Encourage individuals and teams to initiate and participate in improvement activities; offer opportunities (e.g. internal workshops) to share experiences and good practices regarding data management within the organization	EDQM Maturity Model criterion 2b (Ofner, Otto et al., 2013, p. 14) Data quality improvement as a focus area of MDM (Spruit & Pietzka, 2015, p. 1072) "A governance structure […] enables best practice sharing/continuous improvement" (PIC, 2016, p. 9). Organizational capability review, review of personal objectives, data issue management, customer feedback resolution, and process issue management as data quality capabilities (GS1, 2010). Data error cause analysis and data error correction as key DQM processes (ISO, 2011) Data management is a long-term, continuous effort requiring constant adjustment (Lee et al., 2006; Ofner, Otto et al., 2013). PMI case, Schaeffler case, SBB 2 case DXM Maturity Assessment criteria 9.1 and 9.2 (see *Appendix C.1*)	Definition

Activities in the Continuous Improvement design area result in the documentation of a continuous improvement process (see *Table 6-28*).

Table 6-28: Key result document and supporting artifacts of the "Continuous Improvement" design area

Key result document	Supporting artifacts
− Documentation of improvement measures	Data management improvements (Hopkins et al., 2018) Continuous improvement (Bhuiyan & Baghel, 2005; Lindberg & Berger, 1997; Wu & Chen, 2006), Six sigma (Linderman, Schroeder, Zaheer, & Choo, 2003)

7. Reference Model Evaluation

According to the recommendations for ADR given by Sein et al. (2011), evaluation plays an important role in the process of designing an artifact. This chapter presents the results from evaluating the DXM as the central artifact of the research activities described in this dissertation. Following Venable et al.'s (2016) recommendations, the author made use of three distinct evaluation strategies: (1) naturalistic evaluation through application of the artifact in different company settings and for different purposes (*Section 7.1*); (2) formal evaluation including a questionnaire-based survey and an analysis of the extent to which the design requirements are met (*Section 7.2*); and (3) comparison with competing artifacts (*Section 7.3*).

All three evaluation strategies confirm the artifact's general applicability and practical utility, its conformity with the evaluation criteria defined by (Prat, Comyn-Wattiau, & Akoka, 2015), and its compliance with (Becker et al., 1995) "guidelines for orderly modeling". Furthermore, the evaluation provides evidence that the DXM meets all initially defined design requirements, and that this reference model is superior to the competing artifacts introduced in *Chapter 3*.

7.1 Naturalistic Evaluation

As the artifact was adopted by a large number of enterprises inside and outside the research consortium[41], both the validity of the twelve design areas specified by the DXM and the reference model's applicability and practical utility could be proven. The lessons learned from analyzing the ten case studies (see *Chapter* 5) can be summarized as follows:

- The structure of the DXM and the grouping of the design areas into three sections appear to be reasonable and useful, as the case studies cover either all design areas (SBB 2, Schaeffler, tesa), or only those from the Goals and the Results section (Bayer 1), or only those from the Enablers section (Bosch, Bayer 2, SBB 1).

[41] Since the artifact is publicly available on the CC CDQ website (http://www.cc-cdq.ch/data-excellence-model) and as a research report (Pentek & Legner, 2017), and as it forms the basis of an executive training program (see the CDQ Academy example in *Appendix A.4*), it is difficult to estimate the exact number of users.

- The reference model has a generic character. It can be applied to traditional MDM scenarios (Bayer 1 and 2, SBB 1, tesa) and to strategic data management scenarios (PMI, SAP 1 and 2, SBB 2, Schaeffler, Bosch) likewise.
- The reference model is applicable to situational designs, which can be very different (e.g. while PMI and Bayer 2 both address the People, Roles, and Responsibilities design area, the concrete design of their data management organizations differs significantly). This underlines the general applicability of the reference model on the one hand, while at the same time allowing for concrete, company-specific instantiations on the other.

The case studies presented form the empirical basis of this dissertation. Each case study underlines the practical relevance of the research topic. The case studies provide evidence of the applicability and practical utility of the research artifact developed from a practical perspective. The broad adoption of the DXM in each of the seven case studies presented in *Sections 5.4* and 5.5 and in further companies (also without the author being involved) indicate the practical value of the reference model.

Insights gained during the research process, and first instantiations of the artifact in the case studies (see *Chapter 5*), allowed the author to identify typical scenarios for applying the reference model in practice (see *Table 7-1*). These scenarios can be categorized in two groups: (1) translating the abstract design knowledge into a concrete situational design, and (2) using the reference model as abstract situational knowledge for communication, education, maturity assessment, and benchmarking purposes.

Table 7-1: Case studies categorized by application scenario

Scenario		Case study
1. From generic artifact (abstract design knowledge) to concrete situational design (instantiation)	1.1 Reference model applied and adapted for developing concrete situational models	– **Bayer 1: Developing a Data Management Strategy** (*Subsection 5.4.1*) Bayer developed a data strategy based on the alpha version of the reference model. The strategy was detailed by defining activities for improving each design area of the Enablers section of the DXM. – **Bayer 2: Developing a Corporate Directive** (*Subsection 5.4.2*) and **SBB 1: Developing a Governance Policy** (*Subsection 5.4.3*) Both companies developed and issued a corporate directive/policy for data management based on the reference model. These directives/policies rely on the reference model to introduce definitions, outline the design areas, and instantiate each Enabler through relevant standards. – **Bosch: Developing a Data Strategy** *(Subsection 5.4.5)* The company developed a data strategy following the logic of the reference model in terms of explicitly addressing the design areas of the Enablers section. In addition, in the process of developing the data strategy, Bosch took advantage of the Data Strategy Canvas.
	1.2 Reference model applied and adapted for explicating and analyzing (emergent) concrete situational models	– **Schaeffler: Extending Data Management to New Data Domains** (*Subsection 5.4.4*) Schaeffler analyzed its current data management practices for condition-based machine monitoring using the reference model. The goal was to develop a shared understanding of data management in this new domain among different stakeholders.
2. Generic artifact (abstract design knowledge)	2.1 Reference model as a communication and education tool	– **tesa: Communicating Data Management Activities** (*Subsection 5.5.2*) tesa leveraged the reference model to inform its employees about data management and to communicate data management internally. – **CDQ Academy: Structuring a Data Management Training Concept** (*Appendix A.4*) The generic reference model builds the basis for a training program for data management professionals.
	2.2 Reference model as a basis for maturity assessments and benchmarking	– **SBB 2: Developing a Maturity Model** (*Subsection 5.5.1*) SBB regularly assesses the maturity of its data management activities to monitor progress and identify improvements.

7.1.1 Instantiating a Concrete Situational Design

In such a concrete situational design scenario, the reference model is applied in and adapted to an enterprise setting to create a company-specific instantiation. Such an instantiation is either a prescriptive or descriptive situational design (Goldkuhl, 2011; Winter, 2008). Prescriptive instantiations describe to-be scenarios, like a data management strategy (Bayer 1) or a corporate policy (Bayer 2 and SBB 1). Tailoring the generic reference model to the situational model may include company-specific adaptations (e.g. definition of company standards) and may involve renaming of design areas or adjusting the model's layout (i.e. colors or symbols) for compliance with corporate guidelines. Descriptive instantiations explicate an existing or emerging design. With the boundaries of data management being extended to other data domains, the Schaeffler case study exemplifies how the reference model informed a digitalization initiative about existing and required data management capabilities.

7.1.2 Applying the Generic Artifact

The reference model also serves as a communication and education tool. It supports communication – for instance in a company magazine, on a company's intranet, or in a company video – and motivates internal data management activities by referring to an established body of abstract design knowledge (e.g. tesa's data management video, Bayer's wiki). Furthermore, the reference model has been used to structure and develop an education program for data management professionals (i.e. the CDQ Academy example). Finally, the reference model serves as the underlying domain model of a maturity model to assess and benchmark a company's data management activities (as exemplified by the SBB 2 case study).

7.2 Formal Evaluation

7.2.1 Questionnaire-based Evaluation

To evaluate the alpha version of the artifact[42], the author conducted a standardized, questionnaire-based survey among 25 experienced data managers (research activity 1-10)., The questionnaire (see *Appendix C.3*) comprised 24 five-point Likert-scale questions

[42] The alpha version is depicted in *Figure 5-2* in the Bayer 1 case study documentation. It differs from the final version of the artifact in terms of its graphical representation and an additional design area ("Realization", which was deleted in a later design iteration based on the lessons learned from the questionnaire-based evaluation).

and seven open questions following the design artifact evaluation criteria presented by Prat et al. (2015). After a presentation of the reference model, participants were asked to evaluate the reference model's structure (i.e. in terms of completeness, simplicity, clarity, style, homomorphism, level of detail, consistency), adaptability (i.e. robustness, learning capability), and environmental fit (i.e. personal and organizational utility, understandability, organizational fit).

86 percent of the respondents confirmed that the reference model is useful for their data management activities (see *Table C.4* in *Appendix C*[43]). The majority of the survey participants agreed that the reference model covers all relevant areas of data management (88 percent), that it depicts the reality of data management (83 percent), and that it is robust enough to reflect future changes in the data management environment (80 percent). Furthermore, the majority of participants agreed that the reference model provides a sufficient link between data management and business (80 percent) and that it covers the extended scope of data management beyond traditional MDM (85 percent).

Areas with low agreement rates led to modifications of the reference model. For instance, the relatively low score relating to the visualization of the reference model – less than half of the respondents (48 percent) agreed that the graphical layout and design is appropriate – led to a redesign of the model's graphical representation involving professional designers and a focus group in which design options were discussed (research activity 1-12)[44].

7.2.2 Evaluation Against Design Requirements

Evaluating the resulting artifact against initially defined requirements follows (Fettke & Loos, 2003b) recommendation for conducting an analytical evaluation of a reference model. As DSR artifacts are characterized by their practical relevance and scientific rigor (Hevner et al., 2004), the *practical relevance* of the DXM has been assessed by evaluating to which extent the six design requirements introduced in *Table 6-1* (in *Section 6.1*) are met. *Table 7-2* argues that the reference model sufficiently addresses all design requirements concerning the purpose, boundaries, rationale, improvement, and assessment of data management.

Scientific rigor of the DXM is assessed against the "guidelines for orderly modeling" (Becker et al., 1995; Schütte, 1998; Schütte & Rotthowe, 1998), which constitute a

---

[43] Highlighted rows in the table indicate the survey results referred to in this paragraph.

[44] *Appendix A.3* provides an overview how the artifact's visualization evolved over time.

normative framework for a qualitative evaluation of reference models[45]. According to (Becker et al., 1995), reference model design has to follow six general requirements: (1) correctness, (2) relevance, (3) cost effectiveness, (4) clarity, (5) comparability, and (6) systematic structure. *Table 7-3* shows that the DXM has been rigorously designed in accordance with the "guidelines for orderly modeling".

[45] The author deliberately refrained from (again) assessing the reference model against the DSR evaluation criteria by Prat et al. (2015), since the questionnaire-based survey was based on these criteria and low ratings from the survey were addressed in subsequent design activities. Furthermore, the results of the questionnaire-based evaluation are (partially) considered in the evaluation against the "guidelines for orderly modeling".

Table 7-2: Evaluation of the reference model against design requirements

No.	Requirement	Evaluation
R1	Identify business-critical data requirements (*business orientation*).	The reference model contains the Goals section reflecting the business orientation of data management and translating business capabilities into data management capabilities.
R2	Outline key constituents of data management (*key constituents*).	The reference model specifies twelve design areas of strategic data management. According to the questionnaire-based evaluation, the DXM covers all relevant aspects of data management.
R3	Manage data originating from multiple sources and being used for multiple purposes (*scope).*	The reference model includes the "data lifecycle" design area in the Enablers section. This design area seeks to manage all processes regarding the creation, acquisition, storage, maintenance, use, and deletion of data ("from cradle to grave"), defines data objects, and documents data sources, data consumers, and data usage contexts. The Schaeffler case study, dealing with transactional data from multiple machines, indicates that the DXM is useful for structuring scenarios with data from multiple sources being used for multiple purposes.
R4	Address relevant data-related concerns *(purpose beyond data quality).*	The reference model comprises the "data excellence" design area, which refers to data management's impact on data concerning data-related aspects such as data compliance, data security, or data privacy, in addition to data quality.
R5	Demonstrate the business value generated by data and the value contribution of data management to business (*value contribution).*	The reference model comprises "business value" as an outcome of "data excellence". This design area refers to data management's impact on business concerning operational excellence (i.e. cost reduction/efficiency gains), new/enhanced business models (i.e. additional revenue), compliance, and mitigated risks.
R6	Develop data management in stages (*implementation*).	The "continuous improvement" design area of the reference model reflects the consideration of data management as a long-term endeavor that is systematically evolving over time. Furthermore, using the DXM as the domain model for a maturity assessment indicates the evolutionary character of data management inherent in the reference model. The DXM maturity questionnaire supports in assessing the current maturity of data management activities in order to identify weaknesses and define improvement activities.

Table 7-3: Evaluation of the reference model against the requirements of orderly modeling

No.	Requirement	Evaluation
1	Correctness	The meta-model and the definition of each design area and its constructs (see *Subsection 6.2.3*) prove the syntactical correctness of the reference model. Furthermore, completeness and consistency of the model were confirmed in the questionnaire-based evaluation of the DXM. The reference model's semantic correctness was confirmed by the case studies, and by the survey results, which all indicate that the model covers all relevant aspects and depicts the reality of data management.
2	Relevance	The reference model is relevant as it serves explicitly defined usage scenarios: (a) defining a basic terminology of data management; (b) defining and specifying the main design areas and their respective deliverables; (c) sharing knowledge regarding supporting concepts, models, methods, and good practices; and (d) assessing the maturity of data management. The case studies provide evidence that the reference model meets the requirements of these usage scenarios.
3	Cost effectiveness	Applying the reference model yields economic benefit to its users, as it is characterized as a prescriptive and descriptive "blueprint" (i.e. it supports researchers and practitioners in defining, structuring, designing, assessing, and implementing data management activities.) Furthermore, the process of designing the reference model meets the requirement of economic efficiency, as the DXM is based on the RBV, performance management approaches, and the Framework for CDQM.
4	Clarity	The reference model's basic structure, its wording, and its graphical representation were discussed, revised, and amended multiple times during the BIE activities. The survey results and, especially, the CDQ Academy case study underline the clarity of the DXM.
5	Comparability	The reference model is formalized through a meta-model and glossary of terms. This allows comparison of the DXM with other reference models.
6	Systematic structure	The reference model specifies eleven design areas of data management in three sections (Goals, Enablers, and Results), which are interlinked in a continuous improvement cycle (being the twelfth design area). Furthermore, each design area is specified by its success criteria, recommended practices, result documents, and supporting artifacts. Finally, the participants of the questionnaire-based survey confirmed that the DXM has a clear structure.

7.3 Comparison with Competing Artifacts

The comparison of the DXM with the competing artifacts[46] introduced in *Chapter 3* considered the following criteria (see *Table 7-4*): origin (i.e. the parties involved), design process, data excellence dimension (compliance, data privacy, data security, data quality), data in scope (master data, big data), and constituents of the artifact (i.e. the twelve design areas of the DXM). Regarding origin and design process, the DXM appears to be the only reference model for strategic data management which provides a consistent and up-to-date body of design knowledge, while integrating views both from the academic and the practitioners' community. The review also indicates that the DXM is the only artifact which fully addresses all criteria related to data excellence dimensions, data in scope, and constituents. Furthermore, it can be observed that these criteria are of general relevance as they are also reflected by other artifacts (though not all are addressed by a single one). This underlines the validity of the capability reference model for strategic data management.

[46] The review of competing artifacts covers only the artifacts presented in *Chapter 3*, which were developed by researchers, industry consortia, analysts, or standardization bodies. The author excluded reference models from consulting firms or software vendors from this comparison, since these models tend to be single-expert and/or single-case induced.

Table 7-4: Comparison of the reference model with competing artifacts

	Origin	**Design process**	**Data excellence dimension: Compliance**	**Data excellence dimension: Data privacy**	**Data excellence dimension: Data security**	**Data excellence dimension: Data quality**	**Data in scope: Master data**	**Data in scope: Big data**	**Design area: Business capabilities**	**Design area: Data management capabilities**	**Design area: Data strategy**	**Design area: People, roles and responsibili-**	**Design areas: Processes and methods**	**Design area: Data lifecycle**	**Design area: Data applications**	**Design area: Data architecture**	**Design area: Performance management**	**Design area: Data excellence**	**Design area: Business value**	**Design area: Continuous improvement**
Data Excellence Model	Research and industry consortium (15+ Europe-based companies, 3 universities)	Developed using ADR	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Framework for CDQM and Maturity Model for EDQM	Research and industry consortium (30+ Europe-based companies, 2 universities)	Developed using consortium research (multi-lateral, longitudinal DSR process)	-	-	-	X	X	-	-	-	X	X	X	(x)	X	X	X	-	-	(x)
Capability Reference Model for Information Service Management	Research and industry consortium (30+ Europe-based companies, 2 universities)	Developed using consortium research (multi-lateral, longitudinal DSR process)	(x)	X	X	X	X	X	X	X	X	(x)	(x)	X	X	X	(x)	(x)	X	-
Big Data Analytics Capability Model	Research (Delphi study with 51 and 43 respondents, survey with 152 responses)	Literature review, Delphi studies	-	-	-	(x)	X	X	(x)	X	X	X	-	-	X	(x)	X	-	-	-
Big Data Resources Framework	Research (2 studies with 232 and 108 responses from big data analysts and CDOs)	Scale development procedure	-	-	-	-	X	X	-	-	(x)	X	-	-	X	-	-	-	-	(x)
Master Data Management Maturity Model	Research (1 company for evaluation involved)	Design science research	-	(x)	X	X	X	-	-	-	-	X	-	X	-	X	X	(x)	X	X
DAMA-DMBOK Framework	Industry consortium (comprising data management experts in 53 active chapters)	Consensus building in consortium *(details not reported)*	X	X	X	X	X	(x)	(x)	(x)	X	X	X	X	X	X	X	(x)	X	X

	Origin	Design process	Data excellence dimension: Compliance	Data excellence dimension: Data privacy	Data excellence dimension: Data security	Data excellence dimension: Data quality	Data in scope: Master data	Data in scope: Big data	Design area: Business capabilities	Design area: Data management capabilities	Design area: Data strategy	Design area: People, roles and responsibili-	Design areas: Processes and methods	Design area: Data lifecycle	Design area: Data applications	Design area: Data architecture	Design area: Performance management	Design area: Data excellence	Design area: Business value	Design area: Continuous improvement
Data Quality Maturity Model	Industry consortium (U.S.-based governmental agencies)	9 focus groups, survey	X	-	-	X	X	-	-	-	-	X	X	-	-	-	X	-	-	(x)
Data Capability Assessment Model	Industry consortium (200+ companies from the financial sector)	Consensus building in consortium *(details not reported)*	X	X	X	X	X	-	-	-	X	X	X	X	X	X	X	(x)	X	-
IBM Data Governance Council Maturity Model	Industry consortium (52 companies, associations, and governmental agencies, 3 universities)	Consensus building in consortium *(details not reported)*	X	X	X	X	X	(x)	X	X	-	X	X	X	(x)	X	X	(x)	X	-
Data Quality Management System	Standardization body (14 companies and 11 GS1 offices)	*Not reported*	(x)	(x)	X	X	X	X	-	-	X	X	X	X	X	(x)	X	X	(x)	X
Master Data Quality Management Framework	Standardization body (no information on contributors to reference model publicly available)	Consensus building in standardization body *(details not reported)*	-	-	-	X	X	X	-	-	(x)	X	X	X	(x)	X	X	(x)	-	X
Data Management Capability Model	Market analyst (no publicly available information on contributors to reference model)	*Not reported*	-	-	X	-	X	(x)	(x)	X	-	(x)	X	(x)	X	X	-	-	-	-
Enterprise Information Management Maturity Model	Market analyst (no publicly available information on contributors to reference model)	*Not reported*	-	-	-	X	X	X	(x)	(x)	X	X	(x)	X	X	-	X	(x)	X	X

X: fully addressed; (x): partially addressed; -: not addressed

8. Conclusion and Outlook

This chapter summarizes the results of this dissertation and its contribution both to the advancement of the body of scientific knowledge and the broadening of the knowledge base of the practitioners' community. Furthermore, it points to the limitations of the research conducted and proposes actions for future research.

8.1 Summary of Results

The dissertation draws the attention to the changing role of data in business, and what impact this has on data management. While scholars have advocated the idea of data as an economic good since the 1980s, the focus of data management has been on data quality and its improvement so far. In the wake of the megatrend of digitalization and data-driven innovation, data is increasingly becoming an integral part of companies' external value proposition. As data is becoming more and more business critical, existing practices in data management need to be rethought.

As a response to Research Question #1 (asking what the main design areas of strategic data management are, and how these design areas are interrelated), the author presents a capability reference model for strategic data management. The model was systematically developed in a longitudinal, multilateral ADR process. Based on the understanding of data as a strategic resource, the reference model builds on performance management and specifies twelve design areas for data management in three categories (Goals, Enablers, and Results), which are interlinked in a Continuous Improvement cycle. The artifact provides a reference for structuring, reviewing, assessing, communicating, establishing, and improving data management in data-driven enterprises. Being a reusable conceptual model, the DXM *describes* the domain of data management and, at the same time, *prescribes* a blueprint for good design of data management measures (Frank, 2007). By addressing the requirements of data-driven enterprises, the reference model closes a gap in research, which is fragmented into traditional, quality-oriented (master) data management and progressive big data management approaches.

As a response to Research Question #2, asking which concepts, models, and methods are suitable to support the establishment of strategic data management, the author identified and specified a set of success criteria, recommended practices, key result documents, and supporting artifacts. Among these artifacts are the Data Strategy Canvas and a maturity model for assessing strategic data management, which both were developed with and instantiated by member companies of the CC CDQ.

8.2 Contribution to State of the Art

Rooted in the RBV, the suggested reference model conceptualizes data management as goal and outcome-oriented capabilities that contribute to business capabilities and are developed in a continuous improvement cycle. It specifies the required resources and capabilities in twelve distinct design areas. In contrast to competing artifacts, the reference model builds on an ontological foundation, accumulates practitioners' and academic perspectives, provides a consistent and up-to-date body of knowledge, and extends the existing knowledge base over time by integrating "defensive" quality-oriented (master) data management approaches with "offensive" (big) data management approaches. It is rooted in the academic literature and in the body of knowledge developed by the CC CDQ over a period of thirteen years. It externalizes and explicates practical knowledge from subject matter experts (business, IT, and data management) representing more than thirty multinational enterprises as (current or previous) members of the CC CDQ. Developed in a practice-oriented ADR process, the artifact and the process of designing it are relevant for two communities: researchers (from the capability research, IS research, and DSR domain) and practitioners from the data management community.

To the **capability research community**, the reference model presents a RBV on data management, explicating data management capabilities as being contingent on business capabilities. This implies that data management approaches, which have been quite uniform in the past, will increasingly vary between companies pursuing different data-driven business models.

For the **IS research community**, the DXM reveals that big data and advanced analytics do not cause "earth-shaking" shifts in data management, but should rather be considered as extensions of traditional capabilities. This implies that fundamental design principles of data management, like explicating a data strategy or developing data management capabilities through organizational and technical aspects, remain relevant. Consequently, the presented reference model is one of the first attempts to combine traditional with progressive data management approaches.

To the **DSR research community**, the research activities described in this dissertation demonstrate how ADR can be beneficial in a longitudinal, multilateral setting (as the DXM culminates and enhances the research results of the CC CDQ, which has been doing consortium research since 2006). Observing the accumulation of knowledge over time in the different versions of the artifact (i.e. the Framework for CDQM, the Maturity Model for EDQM, and now the DXM) reflects data's increasingly important role in business as well as maturing design knowledge in data management. Furthermore, the three phases of the CC CDQ (i.e. ontology, capability-building, reorientation) mirror stages of knowledge accumulation in a domain of growing maturity. These observations provide an opportunity to derive insights regarding the mechanisms of longitudinal, multilateral knowledge accumulation and evolution in DSR.

For **practitioners**, the reference model provides both descriptive and prescriptive design knowledge in the area of data management. As a descriptive artifact, the DXM synthesizes knowledge for communicating, training, and assessing the maturity of data management activities (i.e. application of the generic artifact). As a prescriptive artifact, the reference model provides design knowledge to build, maintain, and enhance data management capabilities that reflect data's evolving role in business. This allows for configuration and instantiation of the reference model in company-specific settings (i.e. instantiating a concrete situational design). The three exploratory case studies presented in the dissertation illustrate data's changing role and what implications this has on data management, while the other seven case studies exemplify application scenarios of the reference model.

8.3 Study Limitations

The research results presented in this dissertation have some limitations. These limitations are mainly due to the research context, the research design, and a bias resulting from the scope of the research conducted.

First, the CC CDQ, being the research context of the artifact's development, poses some limitations to the research results. The consortium includes large companies with a European origin, mainly headquartered in Germany and Switzerland. As these companies share a common understanding of data management through a typically long-standing membership in the consortium and frequent interactions with each other (i.e. five workshops per year), and as they typically have been using the Framework for CDQM as the predecessor of the DXM, the maturity of their data management activities can be considered relatively high and their mindset towards data management as "synchronized". This might limit the research results' validity to companies with a comparable geographic origin, with a similar understanding of data management, and a high maturity of data management capabilities. Furthermore, only few of the member companies are a data company (from the software or financial industry); instead, most CC CDQ member companies are all dealing with transforming their legacy of physical products, services, and business models to meet the requirements of the digital economy.

Second, qualitative case study research is limited by the selection of cases. The companies selected for instantiating the artifact come from the FMCG, automotive, transportation, and pharmaceutical industry. While the results can be considered valid for these industries, it is likely but not proven that they can be reproduced in other industries. Furthermore, the research design is mainly based on qualitative data; consequently, the research results are analytically valid, but quantitative/statistical validity cannot be claimed.

Third, the CC CDQ is a research consortium focusing on data management. Although neighboring disciplines, such as BI or data analytics, were considered in the research process, most interactions during the BIE stage took place with data management experts. This might have resulted in research results with a bias towards data management at the expense of other scientific disciplines and practical knowledge areas.

8.4 Future Research

Future research opportunities following up on the results presented in this dissertation mainly exist in five areas. First, the research activities conducted have provided the basis for enhancing the consortium research method. While two publications (co-)written by the author of this dissertation reflect on the research process and methods applied (Legner et al., 2020; Pentek & Legner, 2020), the consortium research method could be further developed by exploring contribution principles and knowledge accumulation mechanisms in longitudinal, multilateral DSR/ADR settings.

Second, to maintain acceptance and usage of the dissertation's key artifact among practitioners, the maturity model based on the DXM can be further developed by adopting the recently published, new version of the EFQM Excellence Model (EFQM, 2019).

Third, to overcome the limitation of a data management perspective being focused rather on data foundation than on data monetization (see limitation three in the previous section), the DXM can be amended by a data analytics perspective.

Fourth, extending the research scope to enterprises outside the CC CDQ (especially to pure "data companies" with potentially more advanced data analytics capabilities) will provide promising opportunities to study capability-building mechanisms and contingencies in data management.

Fifth, economic, technological, regulatory, and societal trends should be observed on a permanent basis to reflect their impact on data management and the DXM. Potentially relevant trends to be further explored include the impact of emerging open data platforms, developing microservice offerings, and new approaches of data monetization.

9. References

Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems*, *17*(2), i–xxxii.

Abraham, R., Schneider, J., & Vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, *49*, 424–438.

Adelman, S., Moss, L. T., & Abai, M. (2005). *Data strategy*. Upper Saddle River, NJ: Addison-Wesley.

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, *182*, 113–131.

Alexander, D., & Lyytinen, K. (2017). Organizing Successfully for Big Data to Transform Organizations. *Proceedings of the Twenty-Third Americas Conference on Information Systems, Boston, 2017*.

Amit, R., & Schoemaker, P. J. H. (1993). Strategic assets and organizational rent. *Strategic Management Journal*, *14*(1), 33–46.

Andrews, K. R. (1997). The Concept of Corporate Strategy. In N. J. Foss (Ed.), *Oxford management readers. Resources, firms, and strategies: A reader in the resource-based perspective* (52–59). New York, NY, Oxford, Athens: Oxford University Press.

Atkinson, K., & McGaughey, R. (2006). Accounting for Data: A Shortcoming in Accounting for Intangible Assets. *Academy of Accounting and Financial Studies Journal*, *10*(2), 85–95.

Atul, S., Desale Girish, B., & Patil Swati, P. (2017). Big Data and Big Data Management (BDM) with current Technologies –Review. *International Journal of Engineering Research and Applications*, *07*(04), 42–44.

Azevedo, C. L., Iacob, M. E., Almeida, J. P. A., van Sinderen, M., Pires, L. F., & Guizzardi, G. (2015). Modeling resources and capabilities in enterprise architecture: A well-founded ontology-based proposal for ArchiMate. *Information Systems*, *54*, 235–262.

Back, A., Krogh, G. von, & Enkel, E. (2007). The CC Model as Organizational Design Striving to Combine Relevance and Rigor. *Systemic Practice and Action Research*, *20*(1), 91–103.

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J.L. (2016). Transformational Issues of Big Data and Analytics in Networked Business. *MIS Quarterly*, *40*(4), 807–818.

Baghi, E. (2016). A Capability Reference Model for Establishing Data Quality Controlling (Dissertation). University of St. Gallen, St. Gallen.

Baghi, E., Otto, B., & Österle, H. (2013). Controlling Customer Master Data Quality: Findings from a Case Study. *CONF-IRM 2013 Proceedings*.

Baghi, E., Schlosser, S., Ebner, V. S., Otto, B., & Österle, H. (2014). Toward a Decision Model for Master Data Application Architecture. In *2014 47th Hawaii International Conference on System Sciences* (3827–3836). IEEE.

Ballou, D., Wang, R., Pazer, H., & Tayi, G. K. (1998). Modeling Information Manufacturing Systems to Determine Information Product Quality. *Management Science*, *44*(4), 462–484.

Bardhan, I., Demirkan, H., Kannan, P. K., Kauffman, R. J., & Sougstad, R. (2010). An Interdisciplinary Perspective on IT Services Management and Service Science. *Journal of Management Information Systems*, *26*(4), 13–64.

Bärenfänger, R. (2017). Managing Information Services in the Digital Economy (Dissertation). University of St. Gallen, St. Gallen.

Bärenfänger, R., Leveling, J., & Otto, B. (2016). Linking Service- and Capability-Driven Design – Towards a Framework for Designing Digital Businesses. *MKWI 2016*.

Bärenfänger, R., & Otto, B. (2015). Proposing a Capability Perspective on Digital Business Models. In *17th Conference on Business Informatics (CBI)* (17–25).

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, *17*(1), 99–120.

Baskerville, R., Pries-Heje, J., & Venable, J. (2009). Soft design science methodology. In V. Vaishanvi & S. Purao (Eds.), *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology - DESRIST '09* (p. 1). New York, New York, USA: ACM Press.

Baskerville, R. L., & Wood-Harper, A. T. (1996). A critical perspective on action research as a method for information systems research. *Journal of Information Technology*, *11*(3), 235–246.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, *41*(3), 1–52.

Batini, C., & Scannapieca, M. (2006). *Data quality: Concepts, methodologies and techniques*. *Data-centric systems and applications*. Berlin u.a.: Springer.

Bayer. (2018a). *Annual Report 2017*. Leverkusen.

Bayer. (2018b). *Names, Figures, Facts 2017/2018*. Leverkusen.

Becker, J., Algermissen, L., Delfmann, P., & Knackstedt, R. (2002). Referenzmodellierung. *Das Wirtschaftsstudium (WISU)*, *30*(11), 1392–1395.

Becker, J., Delfmann, P., Dreiling, A., Knackstedt, R., & Kuropka, D. (2004). Configurative process modeling–outlining an approach to increased business process model usability. *Proceedings of the 15th IRMA International Conference*, 1–12.

Becker, J., Rosemann, M., & Schütte, R. (1995). Grundsätze ordnungsmässiger Modellierung. *Wirtschaftsinformatik*, *37*(5), 435–445.

Becker, J., & Delfmann, P. (Eds.). (2007). *Reference Modeling*. Heidelberg: Physica-Verlag HD.

Becker, J., Delfmann, P., & Knackstedt, R. (2004). Konstruktion von Referenzmodellierungssprachen Ein Ordnungsrahmen zur Spezifikation von Adaptionsmechanismen für Informationsmodelle. *WIRTSCHAFTSINFORMATIK*, *46*(4), 251–264.

Becker, J., Delfmann, P., & Knackstedt, R. (2007). Adaptive Reference Modeling: Integrating Configurative and Generic Adaptation Techniques for Information Models. In J. Becker & P. Delfmann (Eds.), *Reference Modeling* (27–58). Heidelberg: Physica-Verlag HD.

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009). Developing Maturity Models for IT Management. *Business & Information Systems Engineering*, *1*(3), 213–222.

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly*, *11*(3), 369.

Bērziša, S., Bravos, G., Gonzalez, T. C., Czubayko, U., España, S., Grabis, J., . . . Zdravkovic, J. (2015). Capability Driven Development: An Approach to Designing Digital Enterprises. *Business & Information Systems Engineering*, *57*(1), 15–25.

Bessant, J., & Caffyn, S. (1997). High-involvement innovation through continuous improvement. *International Journal of Technology Management*, *14*(1), 7.

Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. V. (2013). Digital Business Strategy: Toward a Next Generation of Insights. *MIS Quarterly*, *37*(2), 471–482.

Bharadwaj, A. S. (2000). A Resource-Based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation. *MIS Quarterly*, *24*(1), 169.

Bhuiyan, N., & Baghel, A. (2005). An overview of continuous improvement: From the past to the present. *Management Decision*, *43*(5), 761–771.

Böhmer, M., Dabrowski, A., & Otto, B. (2017). Conceptualizing the Knowledge to Manage and Utilize Data Assets in the Context of Digitization: Case Studies of Multinational Industrial Enterprises. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, *11*(4), 402–411.

Boisot, M., & Canals, A. (2004). Data, information and knowledge: Have we got it right? *Journal of Evolutionary Economics*, *14*(1), 43–67.

Borek, A., Parlikad, A. K., Woodall, P., & Tomasella, M. (2014). A risk based model for quantifying the impact of information quality. *Computers in Industry*, *65*(2), 354–366.

Bosch. (2018). *Bosch Today 2018.*

Bosch. (2019). *Bosch Today 2019.*

Bourdreau, A., & Couillard, G. (1999). Systems Integration and Knowledge Management. *Information Systems Management*, *16*(4), 24–32.

Brancheau, J. C., Schuster, L., & March, S. T. (1989). Building and Implementing an Information Architecture. *ACM SIGMIS Database*, *20*(2), 9–17.

Bücker, I., Hermann, M., Pentek, T., & Otto, B. (2016). Towards a Methodology for Industrie 4.0 Transformation. In W. Abramowicz, R. Alt, & B. Franczyk (Eds.), *Lecture Notes in Business Information Processing. Business Information Systems* (Vol. 255, 209–221). Cham: Springer International Publishing.

Buhl, H. U., Röglinger, M., Moser, F., & Heidemann, J. (2013). Big Data. *Business & Information Systems Engineering*, *5*(2), 65–69.

Cambridge University Press. (2014). Cambridge Advanced Learner's Dictionary & Thesaurus: Citizenship.

Chandler, A. D. (1962). *Strategy and structure: Chapters in the history of the American industrial enterprise*. Washington: Beard Books.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, *36*(4), 1165–1188.

Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: A data management perspective. *Frontiers of Computer Science*, *7*(2), 157–164.

Chen, K., Li, X., & Wang, H. (2015). On the model design of integrated intelligent big data analytics systems. *Industrial Management & Data Systems*, *115*(9), 1666–1682.

Chen, P. P.-S. (1976). The entity-relationship model - toward a unified view of data. *ACM Transactions on Database Systems (TODS)*, *1*(1), 9–36.

Clarke, R. (2016). Big data, big risks. *Information Systems Journal*, *26*(1), 77–90.

CMMI Institute. (2014). *Data Management Maturity (DMM) Model*. Pittsburgh, PA.

CMMI Institute. (2018). *A Rich Heritage*. Retrieved from https://cmmiinstitute.com/company

Collis, D. J. (1994). Research Note: How Valuable are Organizational Capabilities? *Strategic Management Journal*, *15*(S1), 143–152.

DalleMule, L., & Davenport, T. H. (2017). What's Your Data Strategy? *Harvard Business Review*, *95*(3), 112–121.

DAMA. (2008a). DAMA-DMBOK Functional Framework: Version 3.02.

DAMA. (2008b). *The DAMA Dictionary of Data Management* (1. ed.). New Jersey: Technics Publications.

DAMA. (2009). *The DAMA Guide to the Data Management Body of Knowledge: (DAMA-DMBOK guide)* (First edition). Bradley Beach, NJ: Technics Publications LLC.

DAMA. (2011). *The DAMA dictionary of data management* (2nd ed.). Bradley Beach, NJ: Technics Publications.

DAMA. (2017). *DAMA-DMBOK: Data management body of knowledge* (2nd edition). Basking Ridge NJ: Technics Publications.

Davenport, T. H. (1998). Putting the enterprise into the enterprise system. *Harvard Business Review*. (76), 121–131.

Davenport, T. H. (2014). *Big data at work: Dispelling the myths, uncovering the opportunities*. Boston, Mass.: Harvard Business Review Press.

Davenport, T. H., Harris, J. G., Long, D. W. de, & Jacobson, A. L. (2001). Data to Knowledge to Results: Building an Analytic Capability. *California Management Review*, *43*(2), 117–138.

Davenport, T. H., & Patil, D. J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*, *90*(5), 70–76.

Davis, C. K. (2014). Beyond data and analysis. *Communications of the ACM*, *57*(6), 39–41.

Delbaere, M., & Ferreira, R. (2007). Addressing the data aspects of compliance with industry models. *IBM Systems Journal*, *46*(2), 319–334.

Dhar, V., Jarke, M., & Laartz, J. (2014). Big Data. *Business & Information Systems Engineering*, *6*(4), 10–12.

Dierickx, I., & Cool, K. (1989). Asset Stock Accumulation and Sustainability of Competitive Advantage. *Management Science*, *35*(12), 1504–1511.

Dreibelbis, A. (2008). *Enterprise master data management: An SOA approach to managing core information*. *Safari Books Online*. Upper Saddle River, N.J: IBM Press/Pearson plc.

Ebner, V. S. (2014). Entwicklung einer Methode zum Entwurf einer Unternehmensdatenarchitektur (Dissertation). University of St. Gallen, St. Gallen.

EDM Council. (2018). *Data Capability Assessment Model (DCAM) Overview.*

EFQM. (2009). *EFQM excellence model*. *EFQM model 2010*. Brussels: EFQM.

EFQM. (2011). *EFQM Framework for Corporate Data Quality Management: Assessing the Organization's Data Quality Management Capabilities*. Brussels.

EFQM. (2016). *EFQM Framework for CDQM 2016*. Brussels.

EFQM. (2019). *The EFQM Model*. Brussels: EFQM.

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, *14*(4), 532.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory Building from Cases: Opportunities and Challenges. *The Academy of Management Journal*, *50*(1), 25–32.

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: What are they? *Strategic Management Journal*, *21*(10-11), 1105–1121.

English, L. (2003). *Total Information Quality Management – A Complete Methodology for IQ Management.* Brookfield, WI: Information Management and Source-Media, Inc.

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, EU 1995.

General Data Protection Regulation, EU 2016.

Even, A., & Shankaranarayanan, G. (2007). Utility-driven assessment of data quality. *ACM SIGMIS Database*, *38*(2), 75.

Fadler, M., & Legner, C. (2019). Managing Big Data and Analytics: Advanced analytics for more data-driven insights and self-learning systems. Retrieved from https://www.cc-cdq.ch/managing-big-data-and-analytics

Falge, C. (2015). Methode zur Strategieentwicklung für unternehmensweites Datenqualitätsmanagement in globalen Konzernen (Dissertation). University of St. Gallen, St. Gallen.

Falge, C., Otto, B., & Österle, H. (2013). Towards a Strategy Design Method for Corporate Data Quality Management. In *Proceedings of the 11th International Conference on Wirtschaftsinformatik* (801–815).

Ferreira, A., & Otley, D. (2009). The design and use of performance management systems: An extended framework for analysis. *Management Accounting Research*, *20*(4), 263–282.

Fettke, P., & Loos, P. (2003a). Classification of reference models: A methodology and its application. *Information Systems and E-Business Management*, *1*(1), 35–53.

Fettke, P., & Loos, P. (2003b). Multiperspective Evaluation of Reference Models – Towards a Framework. In G. Goos, J. Hartmanis, J. van Leeuwen, M. A. Jeusfeld, & Ó. Pastor (Eds.), *Lecture Notes in Computer Science. Conceptual Modeling for Novel Application Domains* (Vol. 2814, 80–91). Berlin, Heidelberg: Springer Berlin Heidelberg.

Fettke, P., & Loos, P. (2004). Referenzmodellierungsforschung. *WIRTSCHAFTSINFORMATIK*, *46*(5), 331–340.

Fettke, P., & Loos, P. (Eds.). (2007). *Reference Modeling for Business Systems Analysis*: IGI Global.

Fleisch, E., Christ, O., & Dierkes, M. (2005). Die betriebswirtschaftliche Vision des Internets der Dinge. In E. Fleisch & F. Mattern (Eds.), *Das Internet der Dinge* (3–37). Berlin/Heidelberg: Springer-Verlag.

Frank, U. (1994). *Multiperspektivische Unternehmensmodellierung: Theoretischer Hintergrund und Entwurf einer objektorientierten Entwicklungsumgebung*. Zugl.: Marburg, Univ., Habil.-Schr., 1993. *Berichte der Gesellschaft für Mathematik und Datenverarbeitung: Vol. 225*. München, Wien: Oldenbourg.

Frank, U. (2007). Evaluation of Reference Models. In P. Fettke & P. Loos (Eds.), *Reference Modeling for Business Systems Analysis* (118–140). IGI Global.

Fraser, P., Moultrie, J., & Gregory, M. (2002). The use of maturity models/grids as a tool in assessing product development capability. In *IEEE International Engineering Management Conference* (244–249). IEEE.

Gable, G. G. (1994). Integrating case study and survey research methods: An example in information systems. *European Journal of Information Systems*, *3*(2), 112–126.

Gartner. (2014). *The Gartner Enterprise Information Management Framework.*

George, G., Haas, M. R., & Pentland, A. (2014). Big Data and Management. *Academy of Management Journal*, *57*(2), 321–326.

Glaser, B. G. (1965). The Constant Comparative Method of Qualitative Analysis. *Social Problems*, *12*(4), 436–445.

Glaser, B. G., & Strauss, A. L. (2017). *Discovery of Grounded Theory: Strategies for Qualitative Research* (First edition). London: Taylor and Francis.

Goldkuhl, G. (2011). The research practice of practice research: Theorizing and situational inquiry. *Systems, Signs & Actions*, *5*(1), 7–29.

Goodhue, D. L., Quillard, J. A., & Rockart, J. F. (1988). Managing the Data Resource: A Contingency Perspective. *MIS Quarterly*, *12*(3), 373–392.

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

Griffin, J. (2008). The Role of the Chief Data Officer. *Information Management*, *18*(2), 28.

GS1. (2010). *Data Quality Framework: Version 3.0*. Issue 2.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, *53*(8), 1049–1064.

Hagiu, A. (2014). Strategic Decisions for Multisided Platforms. *MIT Sloan Management Review*, *42*(6).

Hansen, H. R., Neumann, G., & Mendling, J. (2015). *Wirtschaftsinformatik: Grundlagen und Anwendungen* (11. völlig neu bearbeitete Auflage). Berlin, Germany, Munich, Germany, Boston, Massachusetts: De Gruyter.

Haug, A., & Stentoft Arlbjørn, J. (2011). Barriers to master data quality. *Journal of Enterprise Information Management*, *24*(3), 288–303.

Helfat, C. E., & Peteraf, M. A. (2003). The dynamic resource-based view: Capability lifecycles. *Strategic Management Journal*, *24*(10), 997–1010.

Henderson, R., & Cockburn, I. (1994). Measuring Competence? Exploring Firm Effects in Pharmaceutical Research. *Strategic Management Journal*, *15*(S1), 63–84.

Henderson, R., & Mitchell, W. (1997). The interactions of organizational and competitive influences on strategy and performance. *Strategic Management Journal*. (18 (Summer Special Issue)), 5–14.

Henriksen, J. P. (2017, September). *Applying the Data Excellence Model at Bayer*. TBM –Master Data Excellence in Life Sciences (3rd Annual Forum), Berlin.

Hermann, M., Pentek, T., & Otto, B. (2016). Design Principles for Industrie 4.0 Scenarios. In *49th Hawaii International Conference on System Sciences (HICSS)* (3928–3937).

Hevner, A. R. (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, *19*(2), 87–92.

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

Hodge, G. M. (2004). *Understanding metadata*. Bethesda MD: NISO Press.

Hofmann, E., & Rüsch, M. (2017). Industry 4.0 and the current status as well as future prospects on logistics. *Computers in Industry*. (89), 23–34.

Hopkins, B., Leganza, G., Goetz, M., & Lee, J. (2018). *Stretch Your Data Management Capabilities: Continuous Improvement: The Data Management Playbook* (No. 87742).

Horlacher, A., & Hess, T. (2016). What Does a Chief Digital Officer Do? Managerial Tasks and Roles of a New C-Level Position in the Context of Digital Transformation. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (5126–5135). IEEE.

Horne, N. W. (1995). Information as an Asset—The Board Agenda. *Computer Audit Update*, *1995*(9), 5–11.

Hüner, K. M. (2011). Führungssysteme und ausgewählte Massnahmen zur Steuerung von Konzerndatenqualität (Dissertation). University of St. Gallen, St. Gallen.

Hüner, K. M., Ofner, M. H., & Otto, B. (2009). Towards a Maturity Model for Corporate Data Quality Management. In D. Shin (Ed.), *Proceedings of the ACM Symposium on Applied Computing.*

Hüner, K. M., Otto, B., & Österle, H. (2011). Collaborative management of business metadata. *International Journal of Information Management*, *31*(4), 366–373.

Hüner, K. M., Otto, B., Österle, H., & Brauer, B. (2011). Fachliches Metadatenmanagement mit einem semantischen Wiki. *HMD Praxis Der Wirtschaftsinformatik*, *48*(1), 98–108.

Hüner, K. M., Schierning, A., Otto, B., & Österle, H. (2011). Product data quality in supply chains: The case of Beiersdorf. *Electronic Markets*, *21*(2), 141–154.

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

Informatica. (2013). *Holistic Data Governance: A Framework for Competitive Advantage.*

International Data Spaces Association. (2019). *Reference Architecture Model: Version 3.0.*

International Organization for Standardization (ISO). (2011). *Data quality — Part 150: Master data: Quality management framework: ISO 8000-150:2011.*

International Organization for Standardization (ISO). (2014a). *ISO 55001:2014: Asset management – Requirements.*

International Organization for Standardization (ISO). (2014b). *ISO 55002:2014: Asset management – Management systems – Guidelines for the application of ISO 55001.*

Jain, H., Ramamurthy, K., Ryu, H.-S., & Yasai-Ardekani, M. (1998). Success of Data Resource Management in Distributed Environments: An Empirical Investigation. *MIS Quarterly*, *22*(1), 1–29.

Joshi, K., & Rai, A. (2000). Impact of the quality of information products on information system users' job satisfaction: An empirical investigation. *Information Systems Journal*, *10*(4), 323–345.

Kagermann, H. (2015). Change Through Digitization—Value Creation in the Age of Industry 4.0. In H. Albach, H. Meffert, A. Pinkwart, & R. Reichwald (Eds.), *Management of Permanent Change* (23–45). Wiesbaden: Springer Fachmedien Wiesbaden.

Kagermann, H., Wahlster, W., & Helbig, J. (2013). Recommendations on implementing the strategic initiative Industrie 4.0: Final report.

Kahn, B., Strong, D., & Wang, R. Y. (2003). Information quality benchmarks: product and service performance. *Communications of the ACM*, *45*(4), 184–192.

Kaplan, R. S., & Norton, D. P. (1996). *The balanced scorecard: Translating strategy into action*. Boston: Harvard Business Review Press.

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, *53*(1), 148.

Kiron, D., & Shockley, R. (2011). Creating Business Value with Analytics. *MIT Sloan Management Review*, *53*(1), 57–63.

Kogut, B., & Zander, U. (1992). Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. *Organization Science*, *3*(3), 383–397.

Kokemüller, J., & Weisbecker, A. (2009). Master Data Management: Products and Research. *ICIQ*, 8–18.

Korte, T., Fadler, M., Spiekermann, M., Legner, C., & Otto, B. (2018). *Data Catalogs: Integrated Platforms for Matching Data Supply and Demand*. Reference Model and Market Analysis.

Krcmar, H. (Ed.). (2015). *Informationsmanagement*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International Journal of Information Management*, *34*(3), 387–394.

Labadie, C., & Legner, C. (2017). *Overview of Data Management Frameworks: Working report.*

Langley, P., & Simon, H. A. (1995). Applications of machine learning and rule induction. *Communications of the ACM*, *38*(11), 54–64.

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

Lee, Y. W., Madnick, Stuart E., Wang, R. Y., Wang, F. L., & Zhang, H. (2014). A Cubic Framework for the Chief Data Officer: Succeeding in a World of Big Data. *MIS Quarterly Executive*, *13*(1), 1–13.

Lee, Y. W., Pipino, L. L., Funk, J. D., & Wang, R. Y. (2006). *Journey to data quality*. Cambridge, Mass.: MIT Press.

Legner, C., Pentek, T., Ofner, M. H., & Labadie, C. (2017). *CDQ Trend Study*. St. Gallen.

Legner, C., Pentek, T., & Otto, B. (2020). Accumulating Design Science Research Knowledge with Reference Models: Insights from 12 Years' Research into Data Management. *Journal of the Association for Information Systems*, *21*(3), 735–770.

Leimeister, J. M. (2015). *Einführung in die Wirtschaftsinformatik*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Leimeister, J. M., Österle, H., & Alter, S. (2014). Digital services for consumers. *Electronic Markets*, *24*(4), 255–258.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing*, *80*(6), 69–96.

Levitin, A. V., & Redman, T. C. (1998). Data as a Resource: Properties, Implications, and Prescriptions. *Sloan Management Review*, *40*, 89–101.

Levy, Y., & Ellis, T. J. (2006). A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. *Informing Science: the International Journal of an Emerging Transdiscipline*, *9*, 181–212.

Lindberg, P., & Berger, A. (1997). Continuous improvement: Design, organisation and management. *International Journal of Technology Management*, *14*(1), 86.

Linderman, K., Schroeder, R. G., Zaheer, S., & Choo, A. S. (2003). Six Sigma: A goal-theoretic perspective. *Journal of Operations Management*, *21*(2), 193–203.

Loshin, D. (2009). *Master data management*. *Safari Books Online*. Amsterdam: Morgan Kaufmann OMG Press/Elsevier.

Lycett, M. (2013). 'Datafication': Making sense of (big) data in a complex world. *European Journal of Information Systems*, *22*(4), 381–386.

Madnick, S. E., Wang, R. Y., Lee, Y. W., & Zhu, H. (2009). Overview and Framework for Data and Information Quality Research. *Journal of Data and Information Quality*, *1*(1), 1–22.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*(4), 251–266.

Marino, K. E. (1996). Developing consensus on firm competencies and capabilities. *Academy of Management Perspectives*, *10*(3), 40–51.

Mata, F. J., Fuerst, W. L., & Barney, J. B. (1995). Technology and Sustained Competitive Advantage: A Resource-Based Analysis. *MIS Quarterly*, *19*(4), 487–506.

Matt, C., Hess, T., & Benlian, A. (2015). Digital Transformation Strategies. *Business & Information Systems Engineering*, *57*(5), 339–343.

Mertens, P., Bodendorf, F., König, W., Schumann, M., Hess, T., & Buxmann, P. (2017). *Grundzüge der Wirtschaftsinformatik*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Mettler, T., Rohner, P., & Winter, R. (2010). Towards a Classification of Maturity Models in Information Systems. In A. D'Atri, M. de Marco, A. M. Braccini, & F. Cabiddu (Eds.), *Management of the Interconnected World* (333–340). Heidelberg: Physica-Verlag HD.

Meuser, M., & Nagel, U. (2009). The Expert Interview and Changes in Knowledge Production. In A. Bogner, B. Littig, & W. Menz (Eds.), *Interviewing Experts* (17–42). London: Palgrave Macmillan UK.

Mieg, H. A., & Brunner, B. (2004). Experteninterviews: Reflexionen zur Methodologie und Erhebungstechnik. *Swiss Journal of Sociology*, *30*(2), 199–222.

Mintzberg, H. (1994). *The rise and fall of strategic planning*. New York: Prentice Hall.

Moingeon, B., Ramanantsoa, B., Métais, E., & Orton, J.D. (1998). Another look at strategy–structure relationships: The resource-based view. *European Management Journal*, *16*(3), 297–305.

Möller, K., Otto, B., & Zechmann, A. (2017). Nutzungsbasierte Datenbewertung. *Controlling*, *29*(5), 57–66.

Möller, K., Wirnsperger, F., & Gackstatter, T. (2015). Performance Management – Konzept, Erfahrungen und Ausgestaltung einer neuen Disziplin. *Controlling*, *27*(2), 74–80.

Moody, D., & Walsh, P. (1999). Measuring The Value Of Information: An Asset Valuation Approach. In *Seventh European Conference on Information Systems (ECIS'99)* (496–512).

Morgan, D. L. (1996). Focus Groups. *Annual Review of Sociology*, *22*(1), 129–152.

Gartner. (2008). *Gartner Introduces the EIM Maturity Model: ID Number: G00160425*. Retrieved from https://www.yumpu.com/en/document/view/24634038/gartner-introduces-the-eim-maturity-model-eurim

Nunamaker, J. J. F., & Briggs, R. O. (2011). Toward a broader vision for Information Systems. *ACM Transactions on Management Information Systems*, *2*(4), 1–12.

O'Brien, R. (2014). Is Traditional Security Dead? *ITNOW*, *56*(1), 26–27.

Ofner, M. H. (2013). Datenqualitätsmanagement aus Prozessperspektive: Methoden und Modelle (Dissertation). University of St. Gallen, St. Gallen.

Ofner, M. H., Hüner, K. M., & Otto, B. (2009). Dealing with Complexity: A Method to Adapt and Implement a Maturity Model for Corporate Data Quality Management. In *Proceedings of the 15th Americas Conference on Information Systems (AMCIS 2009).*

Ofner, M. H., Otto, B., & Österle, H. (2012). Integrating a data quality perspective into business process management. *Business Process Management Journal*, *18*(6), 1036–1067.

Ofner, M. H., Otto, B., & Österle, H. (2013). A Maturity Model for Enterprise Data Quality Management. *Enterprise Modelling and Information Systems Architectures*, *8*(2), 4–24.

Ofner, M. H., Straub, K., Otto, B., & Oesterle, H. (2013). Management of the master data lifecycle: A framework for analysis. *Journal of Enterprise Information Management*, *26*(4), 472–491.

O'Leary, D. E. (2014). Embedding AI and crowdsourcing in the big data lake. *IEEE Intelligent Systems*, *29*(5), 70–73.

OMG. (2017a). *About the Semantics of Business Vocabulary and Business Rules: Version 1.4*.

OMG. (2017b). *Unified Modeling Language: Version 2.5.1*.

Oppenheim, C., Stenson, J., & Wilson, R. M. S. (2003). Studies on Information as an Asset I: Definitions. *Journal of Information Science*, *29*(3), 159–166.

Orr, K. (1998). Data quality and systems theory. *Communications of the ACM*, *41*(2), 66–71.

Österle, H. (2007). Business Engineering - Geschäftsmodelle transformieren. In P. Loos & H. Krcmar (Eds.), *Architekturen und Prozesse* (71–84). Berlin, Heidelberg: Springer Berlin Heidelberg.

Österle, H., Höning, F., & Osl, P. (2011). *Methodenkern des Business Engineering: Ein Lehrbuch*. BE HSG / 2011-1. St. Gallen.

Österle, H., & Otto, B. (2010). Consortium Research: A Method for Researcher-Practitioner Collaboration in Design-Oriented IS Research. *Business & Information Systems Engineering*, *2*(5), 283–293.

Österle, H., & Winter, R. (2003). Business Engineering. In H. Österle & R. Winter (Eds.), *Business Engineering: Vol. 2* (3–18). Berlin: Springer.

Otley, D. (1999). Performance Management: A Framework for Management Control Systems Research. *Management Accounting Research*, *10*, 363–382.

Otto, B. (2011a). Organizing Data Governance:: Findings from the Telecommunications Industry and Consequences for Large Service Providers. *Communications of the Association for Information Systems*, *29*(Article 3), 45–66.

Otto, B. (2011b). Quality Management of Corporate Data Assets. In C.-P. Praeg & D. Spath (Eds.), *Quality Management for IT Services* (193–209). IGI Global.

Otto, B. (2012a). *Enterprise-Wide Data Quality Management in Multinational Corporations*. Cumulative Postdoctoral Thesis.

Otto, B. (2012b). Managing the business benefits of product data management: The case of Festo. *Journal of Enterprise Information Management*, *25*(3), 272–297.

Otto, B., Ebner, V. S., & Hüner, K. M. (2010). Measuring Master Data Quality: Findings from a Case Study. In *Proceedings of the 16th Americas Conference on Information Systems (AMCIS 2010).*

Otto, B., Hüner, K. M., & Österle, H. (2009). Identification of Business Oriented Data Quality Metrics. In P. Bowen, A. K. Elmagarmid, H. Österle, & K.-U. Sattler (Eds.), *Proceedings of the 14th International Conference on Information Quality (ICIQ'09)* (122–134).

Otto, B., Hüner, K. M., & Österle, H. (2010). A Cybernetic View on Data Quality Management. *Proceedings of the 16th Americas Conference*.

Otto, B., Hüner, K. M., & Österle, H. (2012). Toward a functional reference model for master data quality management. *Information Systems and E-Business Management*, *10*(3), 395–425.

Otto, B., Kokemüller, J., Weisbecker, A., & Gizanis, D. (2011). Stammdatenmanagement: Datenqualität für Geschäftsprozesse. *HMD Praxis Der Wirtschaftsinformatik*, *48*(3), 5–16.

Otto, B., & Ofner, M. H. (2010). Towards a Process Reference Model for Information Supply Chain Management. *Proceedings of the European Conference on Information Systems (ECIS 2010)*.

Otto, B., & Österle, H. (2015). *Corporate Data Quality: Prerequisite for Successful Business Models*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Otto, B., & Reichert, A. (2010). Organizing Master Data Management: Findings from an Expert Survey. In S. Y. Shin, S. Ossowski, M. Schumacher, M. J. Palakal, & C.-C. Hung (Eds.), *Proceedings of the 2010 ACM Symposium on Applied Computing* (106–110).

Otto, B., Wende, K., Schmidt, A., & Osl, P. (2007). Towards a Framework for Corporate Data Quality Management. In *Proceedings of the 18th Australasian Conference on Information Systems* (916–926).

Oxford University Press (Ed.). (2004). *Oxford English Dictionary: The Definitive Record of the English Language*. Oxford: Oxford University Press.

Panian, Z. (2010). Some Practical Experiences in Data Governance. *World Academy of Science, Engineering and Technology Management*, 939–946.

Paulk, M. C. (1997). *The Capability maturity model: Guidelines for improving the software process* (6. print). *SEI series in software engineering*. Reading, Mass.: Addison-Wesley.

Paulk, M. C., Weber, C. V., Garcia, S. M., Chrissis, M. B., & Bush, M. (1993). Key Practices of the Capability Maturity Model, Version 1.1.

Pavlou, P. A., & El Sawy, O. A. (2011). Understanding the Elusive Black Box of Dynamic Capabilities. *Decision Sciences*, *42*(1), 239–273.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77.

Pentek, T. (2017). *Data Excellence Model Maturity Assessment: Plenary discussion presentation at the 57th CC CDQ workshop in Cologne.*

Pentek, T., Fadler, M., & Legner, C. (2018). *PMI's Journey Towards a Data-Driven Enterprise: CC CDQ Case Study*. St. Gallen.

Pentek, T., & Legner, C. (2017). *Data Excellence Model: Short Description and Basic Terminology.*

Pentek, T., & Legner, C. (2020). Konsortialforschung zur Entwicklung von Referenzmodellen für die Digitalisierung von Unternehmen: Erfahrungen aus dem Datenmanagement. *HMD Praxis der Wirtschaftsinformatik*, *57*(2), 296–309.

Pentek, T., Legner, C., & Otto, B. (2017). Towards a Reference Model for Data Management in the Digital Economy. In A. Maedche, J. Vom Brocke, & A. R. Hevner (Eds.), *Designing the Digital Transformation: DESRIST 2017 Research in Progress Proceedings of the 12th International Conference on Design Science Research in Information Systems and Technology* (73–82).

Peppard, J. (2003). Managing IT as a Portfolio of Services. *European Management Journal*, *21*(4), 467–483.

PIC. (2016). *Data Quality Maturity Model: PIC Data Quality Working Group Guide.*

Pierce, E. M. (2004). Developing, Implementing and Monitoring an Information Product Quality Strategy. *Proceedings of the Ninth International Conference on Information Quality (ICIQ-04)*, 13–26.

Pinsonneault, A., & Kraemer, K. (1993). Survey Research Methodology in Management Information Systems: An Assessment. *Journal of Management Information Systems*, *10*(2), 75–105.

Pipino, L., Wang, R., Kopcso, D., & Rybolt, W. (2014). Developing Measurement Scales for Data-Quality Dimensions. In X. Wang, E. M. Pierce, S. Madnick, R. C. Fisher, & V. Zwass (Eds.), *Information Quality* (49–63). Hoboken: Taylor and Francis.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data Quality Assessment. *Communications of the ACM*, *45*(4), 211–218.

Porter, M. E. (1980). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

Porter, M. E. (1996). What Is Strategy? *Harvard Business Review*. (November-December).

Porter, M. E., & Heppelmann, J. E. (2015). How Smart, Connected Products Are Transforming Companies. *Harvard Business Review*, *93*(10), 96–114.

Power, D. (2009). Business Process Magement and MDM. *Information Management*, *19*(1).

Prat, N., Comyn-Wattiau, I., & Akoka, J. (2015). A Taxonomy of Evaluation Methods for Information Systems Artifacts. *Journal of Management Information Systems*, *32*(3), 229–267.

Price, R., & Shanks, G. (2005). A semiotic information quality framework: development and comparative analysis. *Journal of Information Technology*. (20), 88–102.

Provost, F., & Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, *1*(1), 51–59.

Redman, T. C. (1995). Improve data quality for competitive advantage. *Sloan Management Review*, *36*(2), 99–107.

Redman, T. C. (1996). *Data quality for the information age*. Boston: Artech House.

Redman, T. C. (1998). The Impact of Poor Data Quality on the Typical Enterprise. *Communications of the ACM*, *41*(2), 79–82.

Redman, T. C. (2001). *Data quality: The field guide*. Boston: Digital Press.

Reichert, A. (2015). Methode zur Einführung von Stammdaten-Management als betriebliche Unterstützungsfunktion (Dissertation). University of St. Gallen, St. Gallen.

Reichert, A., Otto, B., & Österle, H. (2013). A Reference Process Model for Master Data Management. *Proceedings of the 11th International Conference on Wirtschaftsinformatik (WI2013)*, 817–845.

Reid, A., & Catterall, M. (2005). Invisible data quality issues in a CRM implementation. *Journal of Database Marketing & Customer Strategy Management*, *12*(4), 305–314.

Roche, E. M. (1996). Multinational corporations — the emerging research agenda. *The Journal of Strategic Information Systems*, *5*(2), 129–147.

Rosemann, M., & Schütte, R. (1999). Multiperspektivische Referenzmodellierung. In J. Becker, M. Rosemann, & R. Schütte (Eds.), *Referenzmodellierung* (22–44). Heidelberg: Physica-Verlag HD.

Rumelt, R. P. (2011). *Good strategy, bad strategy: The difference and why it matters* (1st ed.). New York: Crown Business.

Sadeghi, A.-R., Wachsmann, C., & Waidner, M. (2015). Security and privacy challenges in industrial internet of things. In *Proceedings of the 52nd Annual Design Automation Conference* (1–6).

Sadiq, S., & Indulska, M. (2017). Open data: Quality over quantity. *International Journal of Information Management*, *37*(3), 150–154.

SAP. (2019). *SAP Corporate Fact Sheet.*

SAS. (2014). *The SAS Data Governance Framework: A Blueprint for Success.*

SBB. (2017a). *Data Strategy 2017-2020*. Bern.

SBB. (2017b). *Governance für das Datenmanagement im Anlagenmanagement Infrastruktur (DM@AM Governance).*

SBB. (2019). *Geschäfts- und Nachhaltigkeitsbericht 2018.*

Schaeffler. (2019). *Annual Report 2018.*

Scheer, A.-W., & Nüttgens, M. (2000). ARIS Architecture and Reference Models for Business Process Management. In W. Aalst, J. Desel, & A. Oberweis (Eds.), *Lecture Notes in Computer Science: Vol. 1806. Business Process Management: Models, Techniques, and Empirical Studies* (376–389). Berlin, Heidelberg: Springer.

Schemm, J. W. (2008). Stammdatenmanagement zwischen Handel und Konsumgüterindustrie: Referenzarchitektur für die überbetriebliche Datensynchronisation (Dissertation). Universität St. Gallen, St. Gallen.

Schlosser, S. (2017). Design principles for collaborative data services (Dissertation). University of St. Gallen, St. Gallen.

Schlosser, S., Baghi, E., Otto, B., & Oesterle, H. (2014). Toward a Functional Reference Model for Business Rules Management. In *2014 47th Hawaii International Conference on System Sciences* (3837–3846). IEEE.

Schmidt, A. (2010). Entwicklung einer Methode zur Stammdatenintegration (Dissertation). University of St. Gallen, St. Gallen.

Schmidt, A., & Otto, B. (2008). A Method for the Identification and Definition of Information Objects. *Proceedings of the 13th International Conference on Information Quality, Cambridge (MA) 2008*, 214–228.

Schüritz, R. M., Seebacher, S., Satzger, G., & Schwarz, L. (2017). Datatization as the Next Frontier of Servitization: Understanding the Challenges for Transforming Organizations. *Proceedings of the Thirty Eighth International Conference on Information Systems, South Korea*.

Schütte, R. (1998). *Grundsätze ordnungsmässiger Referenzmodellierung: Konstruktion konfigurations- und anpassungsorientierter Modelle*. *Neue betriebswirtschaftliche Forschung: Bd. 233*. Wiesbaden: Gabler.

Schütte, R., & Rotthowe, T. (1998). The Guidelines of Modeling – An Approach to Enhance the Quality in Information Models. In G. Goos, J. Hartmanis, J. van Leeuwen, T.-W. Ling, S. Ram, & M. Li Lee (Eds.), *Lecture Notes in Computer Science. Conceptual Modeling – ER '98* (Vol. 1507, 240–254). Berlin, Heidelberg: Springer Berlin Heidelberg.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action Design Research. *MIS Quarterly*, *35*(1), 37–56.

Sen, A. (2004). Metadata management: Past, present and future. *Decision Support Systems*, *37*(1), 151–173.

Senger, E., & Österle, H. (2004). *PROMET Business Engineering Case Studies.*

Setia, P., Setia, P., Venkatesh, V., & Joglekar, S. (2013). Leveraging Digital Technologies: How Information Quality Leads to Localized Capabilities and Customer Service Performance. *MIS Quarterly*, *37*(2), 565–590.

Shankaranarayanan, G., Ziad, M., & Wang, R. Y. (2003). Managing Data Quality in Dynamic Decision Environments: An Information Product Approach. *Journal of Database Management*, *14*(4), 14–32.

Sheng, Y. P. (2003). Exploring the Mediating and Moderating Effects of Information Quality on Firms? Endeavor on Information Systems. In M. Eppler, M. Helfert, L. Pipino, & A. Segev (Eds.), *Proceedings of the 8th International Conference on Information Quality* (344–352).

Sheng, Y. P., & Mykytyn, P. P. (2002). Information Technology Investment and Firm Performance: A Perspective of Data Quality. In C. Fisher & B. N. Davidson (Eds.), *Proceedings of the 7th International Conference on Information Quality* (132–141).

Siau, K., & Rossi, M. (2011). Evaluation techniques for systems analysis and design modelling methods - a review and comparative analysis. *Information Systems Journal*, *21*(3), 249–268.

Smolander, K., Rossi, M., & Pekkola, S. (2017). Infrastructures, Integration and Architecting during and after Digital Transformation. In *2017 IEEE/ACM Joint 5th International Workshop on Software Engineering for Systems-of-Systems and 11th Workshop on Distributed Software Development, Software Ecosystems and Systems-of-Systems (JSOS)* (23–26). IEEE.

Spanos, Y. E., & Lioukas, S. (2001). An examination into the causal logic of rent generation: Contrasting Porter's competitive strategy framework and the resource-based perspective. *Strategic Management Journal*, *22*(10), 907–934.

Spruit, M., & Pietzka, K. (2015). MD3M: The master data management maturity model. *Computers in Human Behavior*, *51*, 1068–1076.

Stahlknecht, P., & Hasenkamp, U. (1999). *Einführung in die Wirtschaftsinformatik* (Neunte, vollständig überarbeitete Auflage). *Springer-Lehrbuch*. Berlin, Heidelberg: Springer.

Stockdale, D. (2014). Data governance and stewardship. In M. S. Brose, M. D. Flood, D. Krishna, & B. Nichols (Eds.), *Handbook of financial data and risk information: Vol. 2. Handbook of financial data and risk information: Software and data* (464–520). Cambridge: Cambridge Univ. Press.

Stoel, D. M., & Muhanna, W. A. (2009). IT capabilities and firm performance: A contingency analysis of the role of industry and IT capability type. *Information & Management*, *46*(3), 181–189.

Suarez, E., Calvo-Mora, A., & Roldán, J. L. (2016). The role of strategic planning in excellence management systems. *European Journal of Operational Research*, *248*(2), 532–542.

Sun, S. X., Zhao, J. L., Nunamaker, J. F., & Sheng, O. R. L. (2006). Formulating the Data-Flow Perspective for Business Process Management. *Information Systems Research*, *17*(4), 374–391.

Susman, G. I., & Evered, R. D. (1978). An Assessment of the Scientific Merits of Action Research. *Administrative Science Quarterly*, *23*(4), 582.

Tan, B., Pan, S. L., Lu, X., & Huang, L. (2015). The Role of IS Capabilities in the Development of Multi-Sided Platforms: The Digital Ecosystem Strategy of Alibaba.com. *Journal of the Association for Information Systems*, *16*(4), 248–280.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509–533.

Tellkamp, C., Angerer, A., Fleisch, E., & Corsten, D. (2004). From Pallet to Shelf: Improving Data Quality in Retail Supply Chains Using RFID. *Cutter IT Journal*, *17*(9), 19–24.

Tesa. (2018). *inside tesa 3/18.*

Tesa. (2019a). *Annual Report 2018*. Norderstedt.

Tesa (2019b, February). *Corporate Data Management @ tesa: Managing data like a true asset*. 63rd CC CDQ Workshop, Düsseldorf, Germany.

Thomas, G. (2014). *The DGI Data Governance Framework.*

Thomas, H., Heene, A., & Sanchez, R. (1996). *Dynamics of competence-based competition: Theory and practice in the new strategic management* (1. ed.). *Technology, innovation, entrepreneurship and competitive strategy series*.

Thomas, O. (2006). Understanding the Term Reference Model in Information Systems Research: History, Literature Analysis and Explanation. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell,. . . A. Haller (Eds.), *Lecture Notes in Computer Science. Business Process Management Workshops* (Vol. 3812, 484–496). Berlin, Heidelberg: Springer Berlin Heidelberg.

Torra, V. (2017). *Data privacy: Foundations, new developments and the big data challenge*. *Studies in big data: Volume 28*. Cham, Switzerland: Springer.

Tremblay, M. C., Hevner, A. R., & Berndt, D. J. (2010). The Use of Focus Groups in Design Science Research. In A. Hevner & S. Chatterjee (Eds.), *Integrated Series in Information Systems. Design Research in Information Systems* (Vol. 22, 121–143). Boston, MA: Springer US.

Van den Hoven, J. (1999). Information Resource Management: Stewards of Data. *Information Systems Management*, *16*(1), 88–90.

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, *25*(1), 77–89.

Vermeer, B. H. (2000). How Important is Data Quality for Evaluating the Impact of EDI on Global Supply Chains? In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences* (1–10).

Vesely, E. G. (1990). *Strategic data management: The key to corporate competitiveness*. *Yourdon Press computing series*. Englewood Cliffs, N.J: Yourdon Press.

Vilminko-Heikkinen, R., & Pekkola, S. (2017). Master data management and its organizational implementation. *Journal of Enterprise Information Management*, *30*(3), 454–475.

Vogel-Heuser, B., Bauernhansl, T., & Hompel, M. ten (Eds.). (2017). *Handbuch Industrie 4.0: Bd. 1: Produktion* (2., erweiterte und bearbeitete Auflage). *Springer Reference Technik*. Berlin: Springer Vieweg.

Vom Brocke, J. (2007). Design Principles for Reference Modeling. In P. Fettke & P. Loos (Eds.), *Reference Modeling for Business Systems Analysis* (47–76). IGI Global.

Vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. *Communications of the*, *37*(1), 205–224.

Wade, M., & Hulland, J. (2004). Review: The Resource-Based View and Information Systems Research: Review, Extension, and Suggestions for Future Research. *MIS Quarterly*, *28*(1), 107–142.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, *34*(2), 77–84.

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an Information System Design Theory for Vigilant EIS. *Information Systems Research*, *3*(1), 36–59.

Wamba, S. F., Akter, S., Trinchera, L., & Bourmont, M. de. (2018). Turning information quality into firm performance in the big data economy. *Management Decision*, *59*(3), 1.

Wand, Y., & Weber, R. (2002). Research Commentary: Information Systems and Conceptual Modeling—A Research Agenda. *Information Systems Research*, *13*(4), 363–376.

Wang, C. L., & Ahmed, P. K. (2007). Dynamic capabilities: A review and research agenda. *International Journal of Management Reviews*, *9*(1), 31–51.

Wang, R. Y. (1998). A Product Perspective on Total Data Quality Management. *Communications of the ACM*, *41*(2), 58–65.

Wang, R. Y., Lee, Y. W., Pipino, L. L., & Strong, D. M. (1998). Manage Your Information as a Product. *Sloan Management Review*, *39*(4), 95–105.

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *JMIS.* Advance online publication.

Ward, J. L., & Daniel, E. (2006). *Benefits Management: Delivering Value from IS and IT Investments*. *John Wiley Series in Information Systems*. Hoboken: John Wiley & Sons Ltd.

Weber, K. (2009). Data Governance-Referenzmodell: Organisatorische Gestaltung des unternehmensweiten Datenqualitätsmanagements (Dissertation). University of St. Gallen, St. Gallen.

Weber, K., Otto, B., & Österle, H. (2009a). Data Governance: Organisationskonzept für das konzernweite Datenqualitätsmanagement. In H. R. Hanse, D. Karagiannis, & H. G. Fill (Eds.), *Proceedings of the 9th International Conference on Wirtschaftsinformatik (WI 2009, Business Services: Konzepte, Technologien, Anwendungen)* (paper 61, 589–598).

Weber, K., Otto, B., & Österle, H. (2009b). One Size Does Not Fit All---A Contingency Approach to Data Governance. *Journal of Data and Information Quality*, *1*(1), 1–27.

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a literature Review. *MIS Quarterly*, *26*(2), xiii–xxiii.

Weill, P., & Woerner, S. L. (2015). Optimizing your digital business model. *MIT Sloan Management Review*, *43*(1), 123–131.

Wende, K. (2007). A Model for Data Governance: Organising Accountabilities for Data Quality Management. In *Proceedings of the 18th Australasian Conference on Information Systems* (417–425).

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018.

Williams, J. (2014). *Internet of Things: Science Fiction or Business Fact (Report and Resources)*.

Winter, R. (2008). Design science research in Europe. *European Journal of Information Systems*, *17*(5), 470–475.

Wixom, B. H., & Ross, J. W. (2017). How to monetize your data. *MIT Sloan Management Review*, *58*(3), 10–13.

Wu, C. W., & Chen, C. L. (2006). An integrated structural model toward successful continuous improvement activity. *Technovation*, *26*(5-6), 697–707.

Wu, S. J., Melnyk, S. A., & Flynn, B. B. (2010). Operational Capabilities: The Secret Ingredient. *Decision Sciences*, *41*(4), 721–754.

Xu, F., Zhan, H., Huang, W., Luo, X. (R.), & Xu, D. (2016). The Value of Chief Data Officer Presence on Firm Performance. *ACIS 2016 Proceeding*, *Paper 213*.

Yin, R. K. (2014). *Case study research: Design and methods* (Fifth edition). Los Angeles: SAGE.

Yuhanna, N., Leganza, G., Karel, R., Evelson, B., Kobielus, J. G., & Owens, L. (2011). *Forrester's Data Management Reference Architecture: Align Data Management Roles To Create Stronger Business Value.*

Zahay, D., & Griffin, A. (2003). Information antecedents of personalisation and customisation in business-to-business service markets. *Journal of Database Marketing & Customer Strategy Management*, *10*(3), 255–271.

Zarnekow, R. (2004). Produktorientiertes Informationsmanagement. In R. Zarnekow (Ed.), *Informationsmanagement: Konzepte und Strategien für die Praxis* (1st ed., 41–56). Heidelberg: dpunkt-Verl.

Zechmann, A. (2016). *Assessing the Economic Value of Data Assets* (Research report). St. Gallen.

Zwitter, A. (2014). Big Data ethics. *Big Data & Society*, *1*(2), 205395171455925.

Glossary

Construct	Definition	Result documents	Justificatory knowledge
Business Capabilities			
External environment	The external environment defines influencing factors outside the enterprise which significantly impact the company's business goals and business operations.	List of external business contingencies	Business capability map (Bärenfänger, 2017, p. 162); Capability modeling (Azevedo et al., 2015) PMI case, Bayer 1 case, Bosch case
Business strategy	A business strategy defines the target state for a company and specifies guiding policies and plans for achieving this state.	Business strategy document	
Business goals	Business goals are derived from the business strategy and define concrete, measurable objectives of strategic relevance for the company.	List of strategic business goals	
Business capabilities	Business capabilities define a set of skills, routines, and resources a company needs to have in order to achieve its business goals through data monetization.	Business capability map	
Data management capabilities			
Data management capability	A data management capability defines a set of skills, routines, and resources a company needs to have in order to accomplish data excellence that results in business value.	Data management capability map	Information service capabilities (Bärenfänger, 2017, p. 156); Reference process model for MDM (Reichert et al., 2013); Functional reference model for CDQM (Otto et al., 2012) Bayer 1 case, Schaeffler case, PMI case
Product and service portfolio	The product and service portfolio of data management defines the data products and data-based services provided by data management to the business.	Portfolio of data products and data management services	Information service portfolio (Bärenfänger, 2017); Portfolio management of IT services (Peppard, 2003; Zarnekow, 2004) Bosch case, PMI case
Data strategy			
Data strategy	A data strategy defines a target state in terms of how data should be managed and used	Data strategy document	

Construct	Definition	Result documents	Justificatory knowledge
	across the entire company, and it develops a plan for achieving this target state.		Strategy Design Method for CDQM (Falge et al., 2013; Falge, 2015); Data management strategy (DAMA, 2017, pp. 32–48); Defensive and offensive data strategies (DalleMule & Davenport, 2017) Bayer 1 case, Bosch case, PMI case, SBB 1 case
Data management goals	Data management goals are derived from the data strategy and define concrete, measurable objectives for the company's data management.	List of data types or domains in scope; list of strategic goals; development plan	
Guiding principles	Guiding principles provide a reference framework for data management activities which reduces the number of options to choose from and provides guidance in daily operations.	Guidelines and principles; code of conduct	
People, Roles, and Responsibilities			
People, roles, and responsibilities	People, roles, and responsibilities define the culture, organization, roles, boards, and interactions for strategic data management.	n/a	Data governance organization layout (Otto & Reichert, 2010; Weber et al., 2009a; Weber, 2009, p. 22); Reference model for data governance (Reichert, 2015; Weber, 2009, p. 106 et seqq.; Weber et al., 2009b); Data governance organization (DAMA, 2017, p. 568) (Redman, 1996, 274ff; Yuhanna et al., 2011); Chief Data Officer (Griffin, 2008; Horlacher & Hess, 2016; Xu et al., 2016); Data Scientist and Data Engineer (Davenport & Patil, 2012) Bosch case, Bayer 2 case, SBB case, tesa case, SAP case 1
Data management organization	A data management organization formally describes the roles and boards involved in data management, the hierarchical layout, and the interactions between these roles and boards.	Organizational structure; list of relevant roles and boards; interaction models	
Data management roles	Data management roles specify the duties, responsibilities, and rights of each data management-related role.	Role profile descriptions; role assignment specifications	
Data management boards	Data management boards specify the duties, responsibilities, and rights of each data management-related board.	Board profile descriptions; list of board members	
Data management responsibilities	Data management responsibilities define required data management-related activities and assign these activities to roles.	RACI matrix	
Corporate culture	A data-driven corporate culture defines an enterprise-wide understanding and awareness of the value and importance of data (as a	Supportive statements; training documents	

Construct	Definition	Result documents	Justificatory knowledge
	strategic resource) at all levels and across all functions and divisions.		
Processes and Methods			
Processes and methods	Processes and methods define procedures and standards for proper and consistent data management.	n/a	Reference process model for master data management (Reichert, 2015, p. 59); MDM process management (Power, 2009) SBB 1 case, Bayer 2 case, Bosch case
Business processes	Business processes define the procedures required to achieve business goals, and specify which tasks are to be executed by whom and in what order.	Business process documentation; list of data (management) requirements	Methodology for data quality-oriented modeling and analysis of business processes (Ofner et al., 2012)
Data management processes	Data management processes define relevant data management procedures on a strategic, governance, and operational level, and specify which tasks are to be executed by whom and in what order.	Data management process documentation	Reference process model for master data management (Reichert, 2015, p. 59); MDM process management (Power, 2009) SBB 1 case, Bayer 2 case, Bosch case
Data management methods	Data management methods specify and illustrate data management activities, with the objective of ensuring standardized, enterprise-wide procedures regarding data management and data use.	Method descriptions; guidelines; standard operating procedures descriptions;	e.g. Strategy Design Method for CDQM (Falge et al., 2013; Falge, 2015); Method for master data integration (Schmidt, 2010, 105 et seqq.); Method for corporate data architecture design (Ebner, 2014, 103 et seqq.); Benefits dependency network (Ward & Daniel, 2006, 133ff)
Data Lifecycle			
Data objects	Data objects represent and define real-world business objects which are generated and/or consumed in business processes.	Data (object) glossary, data management wiki	Method for master data integration: Description of business object types (Schmidt, 2010, p. 105 et seqq.); A Method for the identification and definition of information objects (Schmidt & Otto, 2008); Data knowledge model (Böhmer, Dabrowski, & Otto, 2017)
Business objects	Business objects describe the core entities an enterprise needs to deal with in order to pursue its operations.	Business glossary	

Construct	Definition	Result documents	Justificatory knowledge
			SBB 1 case, Bayer 2 case, Bosch case
Data lifecycles	The data lifecycle manages all processes regarding the creation, acquisition, storage, maintenance, use, archiving, and deletion of data; defines and documents data (objects), data sources, data supply chains, data consumers, and data use contexts.	Data lifecycle documentation; list of business requirements for each data lifecycle stage	Model for information supply chain management (Otto & Ofner, 2010); Management of the master data lifecycle (Ofner, Straub et al., 2013); Reference process model for master data management (Reichert, 2015, p. 59) SBB 1 case, Bayer 2 case, Bosch case
Data lifecycle processes	Data lifecycle processes define the tasks along the data lifecycle and specify which task is to be executed by whom and in what order.	Data lifecycle process documentation	
Data sources	Data sources define internal or external providers of data.	List of data sources used and available; documentation of data sources	Method for master data integration: semantic data model development (Schmidt, 2010, p. 140 et seqq.); Data knowledge model (Böhmer et al., 2017)
Data consumers	A data consumer collects and transforms data from multiple sources to create data products and/or consumes data products and/or data management services.	Documentation of data users and their requirements; documentation of data use contexts	
Data Applications			
Data applications	Data applications is about planning, implementing, and maintaining software which is designed to manage data and data products to achieve and maintain data excellence.	Application landscape documentation; application description	Method for master data integration: semantic data model development (Schmidt, 2010, p. 140 et seqq.); Method for corporate data architecture design (Ebner, 2014, 103 et seqq.); Decision model for master data application architecture (Baghi et al., 2014); Application and integration architecture design (Österle et al., 2011, p. 82 et seqq.); Data integration and interoperability (DAMA, 2017, p. 269 et seqq.) Bosch case, Bayer 2 case, PMI case
Interfaces	Interfaces specify which other applications provide/receive which data to/from the applications documented.	Application interface documentation	
Storage databases	Storage databases define the data repository containing the data required for executing the applications' discrete functions.	Application storage documentation	

Construct	Definition	Result documents	Justificatory knowledge
Functions	Functions document what tasks and processes an application is able to execute.	Functional description of each application	Functional reference model for CDQM (Otto et al., 2012) Method for master data integration: application analysis (Schmidt, 2010, p. 147 et seq.)
Data Architecture			
Data architecture	The data architecture defines and maintains specifications that provide a shared business vocabulary, express strategic data requirements, and outline high-level, integrated architecture landscape designs and data flows.	n/a	Method for master data integration: semantic data model development (Schmidt, 2010, p. 105 et seqq.); A Method for the identification and definition of information objects (Schmidt & Otto, 2008); Method for corporate data architecture design (Ebner, 2014, 103 et seqq.); Meta data management in a semantic wiki (Hüner, Otto, Österle et al., 2011); Collaborative management of business metadata (Hüner, Otto, & Österle, 2011); Implementation of the business vocabulary & rulebook (Schlosser, 2017, p. 141 et seqq.); Data knowledge model (Böhmer et al., 2017); Strategic data planning method (Goodhue et al., 1988); Data modeling and design (DAMA, 2017, p. 123 et seqq.) SBB 1 case, PMI case, Bayer 2 case
Data model	A data model depicts core business objects and the relations between them.	Data model documentation	
Conceptual data model	A conceptual data model depicts the core data objects, the relations between them, and the data requirements from a business perspective.	Documentation of conceptual data model; data (object) glossary/dictionary; data management wiki; metadata repository	
Physical data model	A physical data model depicts the core data objects, the relations between them, and their technical representation from an application perspective.	Documentation of physical data model	Method for master data integration: master data integration architecture design (Schmidt, 2010, p. 105 et seqq.); Method for corporate data architecture design (Ebner, 2014, 103 et seqq.);
Data storage and distribution architecture	The data storage and distribution architecture defines the leading and the consuming	Documentation of leading and consuming	

Construct	Definition	Result documents	Justificatory knowledge
	applications per data object and data attribute.	applications (per data object and data attribute); documentation of data distribution and storage architecture	Decision model for master data application architecture (Baghi et al., 2014); Data storage (DAMA, 2017, p. 169 et seqq.) Bayer 2 case, PMI case
Data flows	Data flows define how and in what order data objects and attributes are exchanged between applications.	Data flow diagram; data lineage documentation	
Business rules	Data management-related business rules are directives that govern, guide or influence the habits and procedures of business departments with regard to data management.	Business vocabulary and rulebook	Business rules management (Schlosser et al., 2014; Schlosser, 2017); Business rules specification and vocabulary (OMG, 2017a) PMI case
Performance Management			
Performance management	Performance management plans, implements and controls all activities for measuring, assessing, improving and ensuring data management performance, data excellence, and business value.	n/a	Capability reference model for data quality controlling (Baghi, 2016); Data quality controlling systems (Hüner, 2011); Measuring data quality (Otto, Ebner et al., 2010)
Performance management system	A performance management system defines metrics, target values, and controlling measures to review and manage the outcome, performance, and maturity of data management.	Metric system; performance management processes	
Data excellence metrics	Data excellence metrics define how data excellence – as an outcome of data management – is measured.	Data excellence metrics	Data quality management (Batini et al., 2009; Batini & Scannapieca, 2006; DAMA, 2017, p. 449 et seqq.; Lee et al., 2006) Schaeffler case, SAP 2 case
Business value metrics	Business value metrics define how data management's contribution to the business is measured.	Business value metrics	Balanced score card (Kaplan & Norton, 1996); Usage-based data valuation (Möller et al., 2017) PMI case, Schaeffler case

Construct	Definition	Result documents	Justificatory knowledge
Data management performance metrics	Data management performance metrics define how the performance of data management activities is measured.	Data management activity metrics, Service Level Agreements	Maturity model for CDQM (Hüner et al., 2009; Ofner, Hüner, & Otto, 2009; Ofner, Otto et al., 2013); Data management performance metrics (Legner et al., 2017, 40 et seq.); Data quality management metrics (DAMA, 2017, p. 449 et seqq.) Schaeffler case
Data management progress metrics	Data management progress metrics define how the progress of data management activities is measured.	Progress metric	Data management progress metrics (Legner et al., 2017, 42 et seq.) PMI case, SBB case, Schaeffler case
Data Excellence			
Data excellence	Data excellence refers to the impact of data management on data quality, but also on additional data-related dimensions, such as regulatory compliance, data security, or data privacy.	Data excellence report; data excellence dashboard; data excellence alert	Data excellence definition (Pentek et al., 2017)
Data excellence dimensions	Data excellence dimensions define the relevant aspects of data management results.	Documentation of relevant data excellence dimensions	
Data quality	Data quality aims at ensuring data's "fitness for purpose" as a context-specific, multi-faceted concept comprising various dimensions, such as consistency, completeness, accuracy, or timeliness.	Data quality report; data quality dashboard; data quality alert	Method for specifying business-oriented data quality metrics (Hüner, 2011, p. 83 et seqq.); Schaeffler case, SAP 2 case Product data quality in supply chains (Hüner, Schierning et al., 2011); Data quality management (Batini et al., 2009; Batini & Scannapieca, 2006; DAMA, 2017, p. 449 et seqq.; Lee et al., 2006)
Data security	Data security aims at protecting data from unauthorized access to meet regulations,	Data security report; data security dashboard;	Data security (DAMA, 2017, p. 217 et seqq.)

Construct	Definition	Result documents	Justificatory knowledge
	contractual agreements, or business requirements.	data security alert	
Data privacy	Data privacy aims at considering general data protection laws and individuals' data privacy preferences when handling personal and other sensitive data.	Data privacy report; data privacy dashboard; data privacy alert	Data privacy aspects (Delbaere & Ferreira, 2007; Sadeghi et al., 2015; Torra, 2017); Data handling ethics (DAMA, 2017, p. 49 et seqq.)
Regulatory compliance	Regulatory compliance aims at meeting the requirements of data-related laws and/or regulations.	Compliance report; compliance dashboard; compliance alert	Data compliance (Delbaere & Ferreira, 2007)
Business Value			
Business value	Business value refers to data's and data management's impact on business.	Business value report; business value dashboard	Benefits dependency network (Ward & Daniel, 2006, p. 133 et seqq.); Balanced score card (Kaplan & Norton, 1996)
Business value dimensions	Business value dimensions define business-related results of data management activities.	Documentation of relevant business value dimensions	
Monetary value	The monetary value refers to the cost, price, or economic value of data.	Data valuation report	Usage-based data valuation (Möller et al., 2017); Data monetization (Wixom & Ross, 2017) PMI case, Bosch case
Business process improvement	Business process improvement aims at utilizing data for decision-making support or increasing enterprise performance.	Business process improvement report; business process improvement dashboard	A cybernetic view on data quality management (Otto, Hüner, & Österle, 2010) Schaeffler case
Customer satisfaction	Customer satisfaction aims at providing excellent data products and/or data management services to business units and/or functions (as internal customers), and at providing excellent data, data products, and/or data-enabled services to external customers and/or suppliers.	Customer satisfaction survey	Datatization (Schüritz et al., 2017); Data monetization (Wixom & Ross, 2017)
Organizational learning and growth	Organizational learning and growth aim at utilizing data to create novel forms of value proposition and data-driven innovation.	Innovation index; report on data-driven revenue or profit generated	
Continuous Improvement			

Construct	Definition	Result documents	Justificatory knowledge
Continuous improvement	Continuous improvement allows adjusting Goals and Enablers of data management based on the Results achieved, ensuring the dynamic nature of data management.	n/a	Continuous improvement (Bhuiyan & Baghel, 2005; Lindberg & Berger, 1997; Wu & Chen, 2006); Six sigma (Linderman et al., 2003); Data management improvements (Hopkins et al., 2018); Data quality improvements (Batini et al., 2009) PMI case, Schaeffler case, SBB 2 case
Continuous improvement process	A continuous improvement process defines routines for initiating improvements as well as techniques to document and control measures.	Documentation of continuous improvement process	
Continuous improvement measures	Measures for continuous improvement define the activities, their documentation, and the monitoring to adapt data management Goals and Enablers based on lessons learned from previous Results.	Documentation of measures for continuous improvement	

Appendix A: Research Context, Methods, and Results

A.1 Presentations Related to the Dissertation

Conference	Topic	Location and Date
48th CC CDQ Workshop	Towards the CDQ Framework 2.0: Motivation, Approach, Pre-considerations, and Plenary Discussion	Hamburg, Germany February 25th, 2016
TBM's 2nd Annual Master Data Excellence Forum in Life Sciences	Trends in Data Management: Corporate Data Valuation	Frankfurt, Germany September 8th, 2016
51st CC CDQ Workshop	CDQ Trend Study 2016 - What data managers need to know	Gottlieben, Switzerland September 14th, 2016
ThinkLinkers MDM Forum	Trends in Data Management: The Financial Value of Data	Berlin, Germany March 27th, 2017
DESRIST 2017	Towards a Reference Model for Data Management for the Digital Economy	Karlsruhe, Germany May 31st, 2017
TBM's 3rd Annual Master Data Excellence Forum in Life Sciences	Applying the Data Excellence Model at Bayer	Berlin, Germany September 14th, 2017
ThinkLinkers MDM ThinkLab	Principles of Future Data Management	Berlin, Germany April 12th, 2018
ECIS 2018: JAIS Paper Development Workshop	A Reference Model for Data Management in the Digital and Data-driven Enterprise	Portsmouth, Great Britain June 26th, 2018
SAP-Infotage Business Data und Analytics	Data Strategy: Increase the Value of Data	Essen, Germany September 12th, 2019
DSAG Thementag Data Strategy	Data Strategy: Increase the Value of Data	Berlin, Germany November 21st, 2019

A.2 Focus Groups and Expert Interviews

Research activity	Location and date	Topic	Participants
1-1	Dortmund, Germany November 4th, 2015	Focus group: CC CDQ Steering Committee 2015: Research requirements	- Dr. Joachim Maasz (Bayer) - Gerhard Gripp (Bayer) - Sabine Kösling-Guse (Bosch) - Eva Schultze (Dräger) - Jerone Walters (Merck) - Philippe Gerwill (Novartis) - Conny Dethloff (Otto) - Andreas Döhrn (SAP) - Markus Rahm (Schaeffler) - Anna Gleiss (Siemens) - Stig Persson (Ericsson) - Karsten Muthreich (Nestlé) - Roger Kipfer (Swisscom) - Christoph Meier (Swisscom) - Timo Neumann (ZF) - Andreas Vogt (ZF) - Henning Uiterwyk (eclass) - Jan-Olav Boeriis (ABB) - Jürgen Jost (Merck) - Andreas Schierning (Beiersdorf) - Dominic Moser (SBB) - Oliver Soine (Bosch)
1-3	Lucerne, Switzerland April 29th, 2016	Focus group: CDQ Framework 2.0	- Christian Geiseler (SAP) - Andreas Schierning (Beiersdorf) - Eva Schultze (Dräger) - Alexander Schmidt (SBB) - Philipp Windmüller (Bayer) - Kjell Korsmo (ABB) - Klaus Pfreundner (Bosch)
1-4	Telephone / video conferences May – August 2016	Opinion survey, expert interview	- Johannes John (Merck) - Mareike Bollinger (Karl Storz) - Karsten Muthreich (Nestlé) - Jürg Hofer (Emmi) - Christian Eck (eclass) - Ben Hallez (Bayer) - Jan-Olav Boeriis (ABB) - Henning Möller (ZF) - Timo Neumann (ZF) - Brigitte Lupp (Osram) - Guillermo Spitzner (Osram) - Alexander Schmidt (SBB) - Christian Trachsel (SBB) - Anna Gleiss (Siemens) - Michael Ericsson (Ericsson) - Göran Domeji (Ericsson) - Matthias Dod (Bosch) - Andreas Schierning (Beiersdorf) - Josef Huber (Festo) - Janika Janssen (Otto) - Eva Schultze (Dräger)

Research activity	Location and date	Topic	Participants
			- Christian Geiseler (SAP)
1-5	Nuremberg, Germany June 29th, 2016	Focus group: CDQ Framework 2.0	- Philip Windmüller (Bayer) - Andreas Lehmann (Festo) - Hennig Möller (ZF) - Markus Rahm (Schaeffler) - Timo Neumann (ZF) - Klaus Pfreundner (Bosch) - Achim Gooren (Schaeffler) - Mareike Bollinger (Karl Storz) - Andrea Brauckhoff (Merck) - Andreas Schierning (Beiersdorf) - Robin Palit (Siemens) - Meinolf Lamberts (Otto)
1-7	Gottlieben, Switzer-land September 14th, 2016	Focus group: CDQ Framework 2.0	- Ian MacLellan (ABB) - Andreas Schierning (Beiersdorf) - Andreas Lehmann (Festo) - Karsten Muthreich (Nestle) - Anne Smorthit (Pilatus) - Hennig Möller (ZF) - Timo Neumann (ZF) - Bernd Schmidt (ZF)
1-8	Ostfildern, Germany November 10th, 2016	Focus group: CC CDQ Steering Committee 2016: Re-search requirements	- Mareike Bollinger (Karl Storz) - Conny Dethloff (Otto) - Josef Huber (Festo) - Jani Karonen (ABB) - Jürgen Jost (Merck) - Klaus Pfreundner (Bosch) - Thorsten Kroke (eclass) - Andreas Lehmann (Festo) - Oliver Soine (Bosch) - Markus Rahm (Schaeffler) - Andreas Schierning (Beiersdorf) - Stefan Schwing (Bayer) - Philippe Baumlin (BASF) - Thomas Gering (RUAG) - Stefan Peetz (LIDL) - Eva Schultze (Dräger)
1-9	Cologne, Germany December 9th, 2016	Focus group: CDQ Framework 2.0	- Tobias Nilsson (Ericsson) - Klaus Pfreundner (Bosch) - Jochen Kokemüller (Bosch) - Jens Peter Henriksen (Bayer) - Jan Hinsch (tesa) - Thomas Affentranger (SBB) - Andreas Schierning (Beiersdorf) - Markus Rahm (Schaeffler) - Philip Windmüller (Bayer) - Mareike Bollinger (Karl Storz) - Andreas Lehmann (Festo) - Hennig Möller (ZF) - Timo Neumann (ZF) - Christian Eck (eclass) - Thorsten Kroke(eclass)

Research activity	Location and date	Topic	Participants
1-12	Berlin, Germany February 22nd, 2017	Focus group: Data Excellence Model	- Kjell Korsmo (ABB) - Marco Pinheiro (AstraZeneca) - Marcus Kottmann (BASF) - Henning Dicke (Bayer) - Andreas Schierning (Beiersdorf) - Roland Schmid (Biogen) - Andreas Lehmann (Festo) - Sandra Roth (Merck) - Achim Gooren (Schaeffler) - Timo Neumann (ZF)
2-4	Bern, Switzerland September 4th, 2017	Expert interview: Maturity Assessment	- Martin Böhlen (SBB) - Alexander Schmidt (SBB) - Peter Kolbe (SBB) - Christian Trachsel (SBB) - Thomas Affentranger (SBB) - Nicolas Casarin (SBB) - Monika Huber (SBB)
2-7	Cologne, Germany December 8th, 2017	Focus group: Data Excellence Model - Maturity Assessment	- Philippe Baumlin (BASF) - Thomas Derichs (Bayer) - Ralf Berges (Bayer) - Andreas Schierning (CDQ Award Jury) - Thomas Bode (Swarovski) - Andreas Lehmann (Festo) - Johannes Sauter (Fraunhofer IAO) - Jürgen Lay (Geberit) - Ramon Palau (RUAG) - Nicolas Casarin (SBB) - Robin Palit (Siemens) - Markus Klein (TÜV Rheinland) - Olaf Albrecht (Webasto)
3-1	Hamburg, Germany April 25th, 2018	Focus group: Data Excellence Model - Handbook and Extensions	- Markus Rahm (Schaeffler) - Robin Palit (Siemens) - Heiko Frank (Bayer) - Marcus Kottmann (BASF) - Jens Eckrich (BASF) - Kantilal Bhimani (PMI) - Armando Bonilla (PMI) - Jean-Philipp Roulet (Helvetia)
3-3	Ludwigshafen, Germany June 22nd, 2018	Focus group: Data (Management) Strategy	- Marcus Kottmann (BASF) - Philippe Baumlin (BASF) - Markus Rahm (Schaeffler) - Danny Thurm (Schaeffler) - Karsten Muthreich (Nestlé) - Matthias Ulrich (Siemens)
3-6	Hamburg, Germany May 9th, 2019	Focus group: Rethinking Data Strategies	- Cornelia Schaeffer (MSD) - Karen Gärtner (Schaeffler) - Matthias Dod (Bosch) - Alexander Gerhard (Radeberger) - Barbara Lienert (SchwarzIT) - Leonie Frank(Swarovski)

Research activity	Location and date	Topic	Participants
			- Michael Ragnieri (Otto) - Jan Hinsch (tesa) - Lianne Bodenstaff (AkzoNobel) - Thomas Rupprecht (Beiersdorf)
3-7	Zurich, Switzerland June 26th, 2019	Focus group: Rethinking Data Strategies	- Matthias Dod (Bosch) - Klaus Pfreundner (Bosch) - Christian Fischer (DB Schenker) - Florian Leser (ADAC) - Martin Böhlen (SBB) - Andreas Lehmann (Festo) - Markus Rahm (Schaeffler)
3-8	August to October 2019	Expert interviews: CDQ Data Strategy Study	- Susan Wegner (Telekom) - Dirk Rahmhorst (Wacker) - Jeff DeWolff, Balaji Rajamani (TetraPak) - Matthias Dod, Albert Hatz (Bosch) - Stefan Peetz (SchwarzIT) - Thomas Bode (Swarovski) - Peter Kolbe (SBB) - Thomas Rupprecht, Alexander Ramin (Beiersdorf) - Jan Hinsch, Robert Yung (tesa) - Sandra Huerlimann, Guiseppe Fragnelli (Helvetia) - Anne Catherine Grögaard (Astra-Zeneca) - Andreas Lehmann (Festo) - Markus Rahm, Jürgen Henn (Schaeffler) - Tina Rosario (SAP) - Rene Zimmermann (Merck) - Stefan Schwing (Bayer) - Carsten Zehnich (ZF) - Rainer Zeusche (Wagner)
3-10	Lausanne, Switzerland October 16th, 2019	Focus group: CC Research Team Discussion of the DXM	- Christine Legner (CC CDQ) - Markus Eurich (CC CDQ) - Martin Fadler (CC CDQ) - Clement Labadie (CC CDQ) - Pavel Krasikov (CC CDQ)
3-12	Ovronnaz, Switzerland January 20th, 2020	Focus group: CC Research Team Discussion of the DXM	- Christine Legner (CC CDQ) - Markus Eurich (CC CDQ) - Martin Fadler (CC CDQ) - Clement Labadie (CC CDQ) - Pavel Krasikov (CC CDQ)

A.3 Evolution of the Artifact Visualization

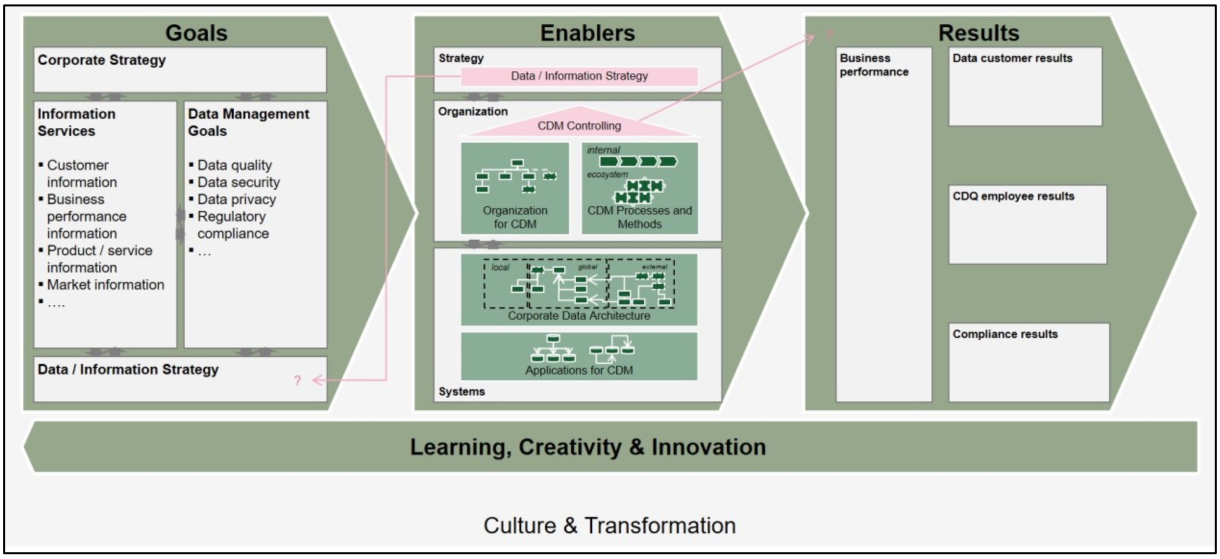


Figure A-1: Reference model presented in research activity 1-2 (plenary discussion in Hamburg, Germany, February 25th, 2016)

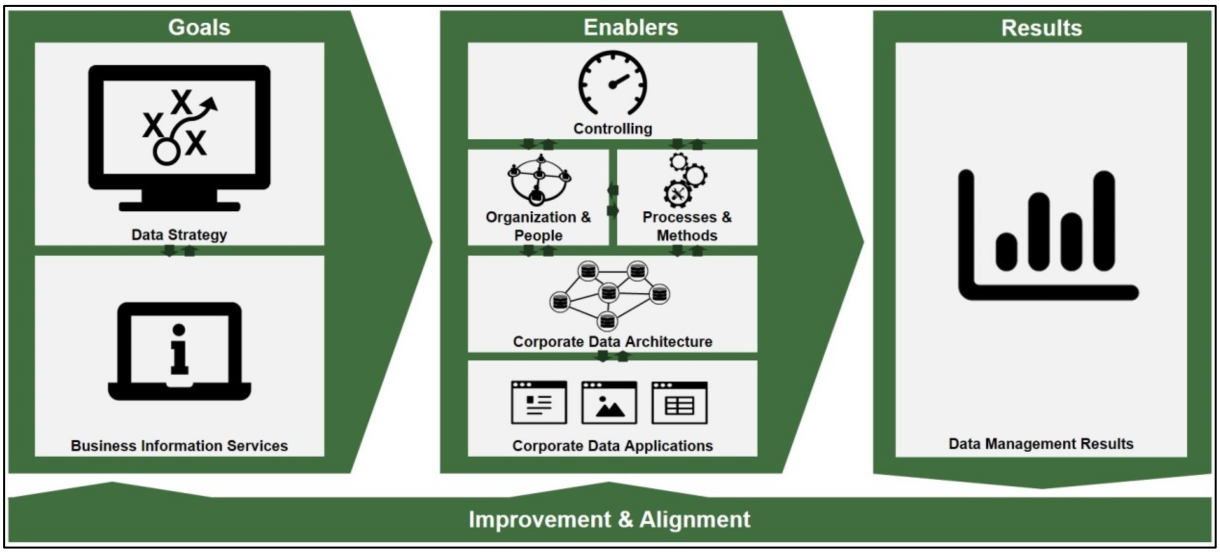


Figure A-2: Reference model presented in research activity 1-5 (focus group in Nuremberg, Germany, June 29th, 2016)

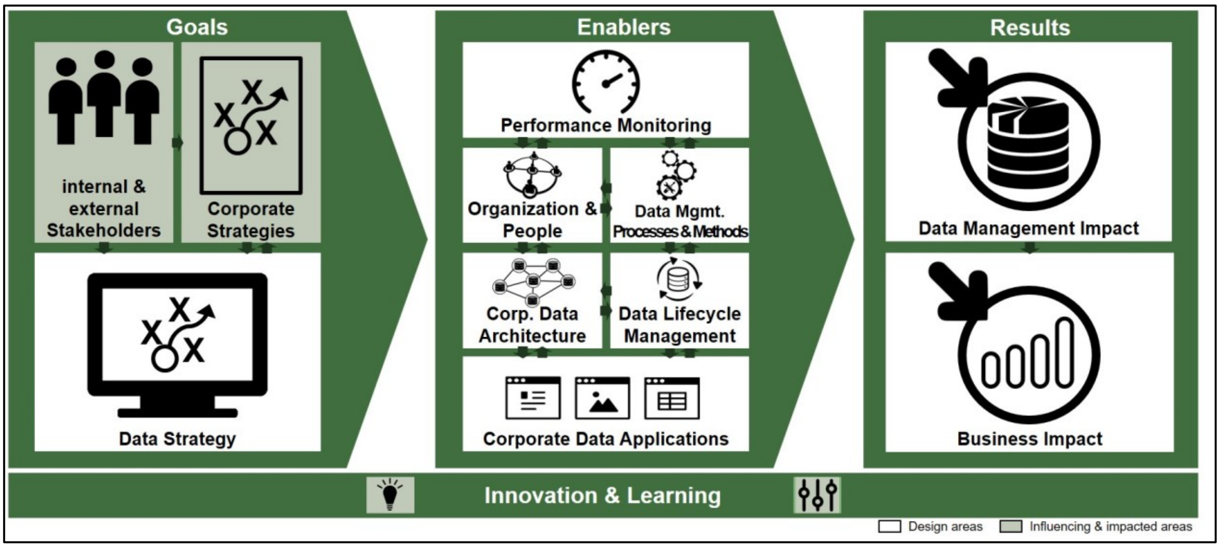


Figure A-3: Reference model presented in research activity 1-7 (focus group in Gottlieben, Switzerland, September 14th, 2016)

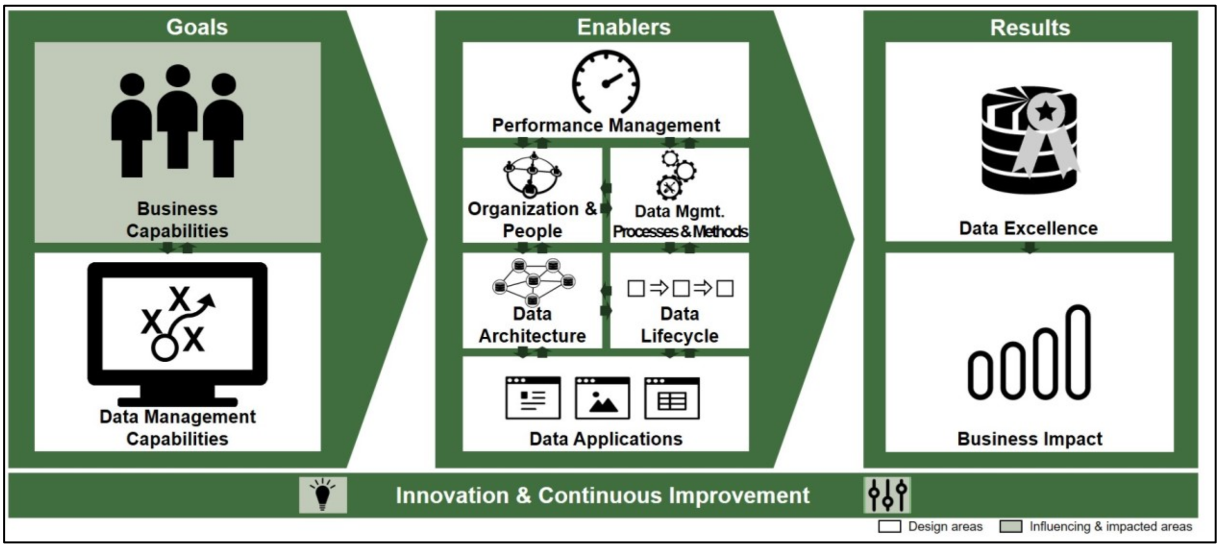


Figure A-4: Reference model resulting from research activity 1-7 (focus group in Gottlieben, Switzerland, September 14th, 2016)

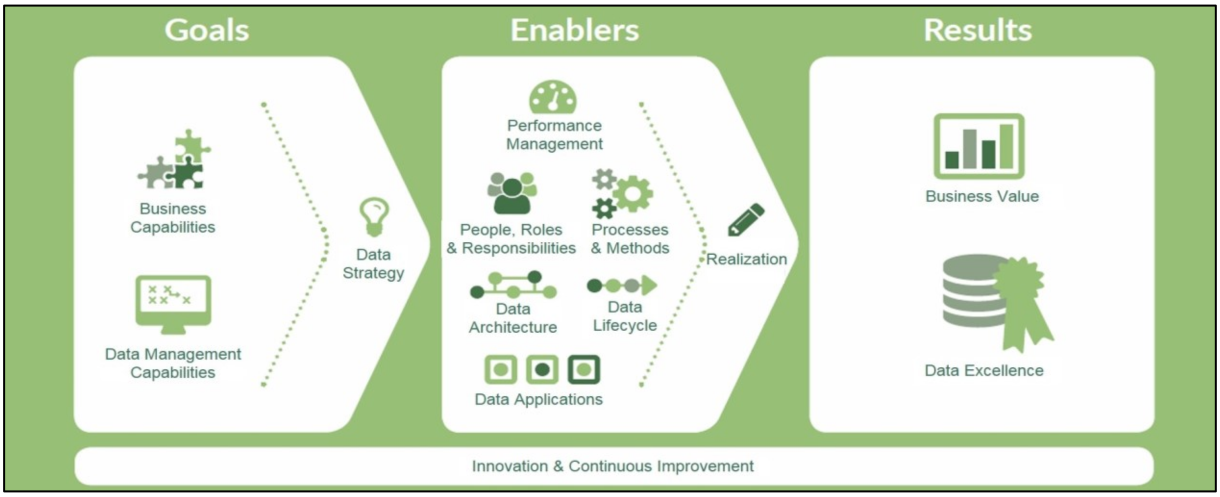


Figure A-5: Reference model presented in research activity 1-9 and evaluated in research activity 1-10 (focus group and opinion survey in Cologne, Germany, December 9th, 2016)

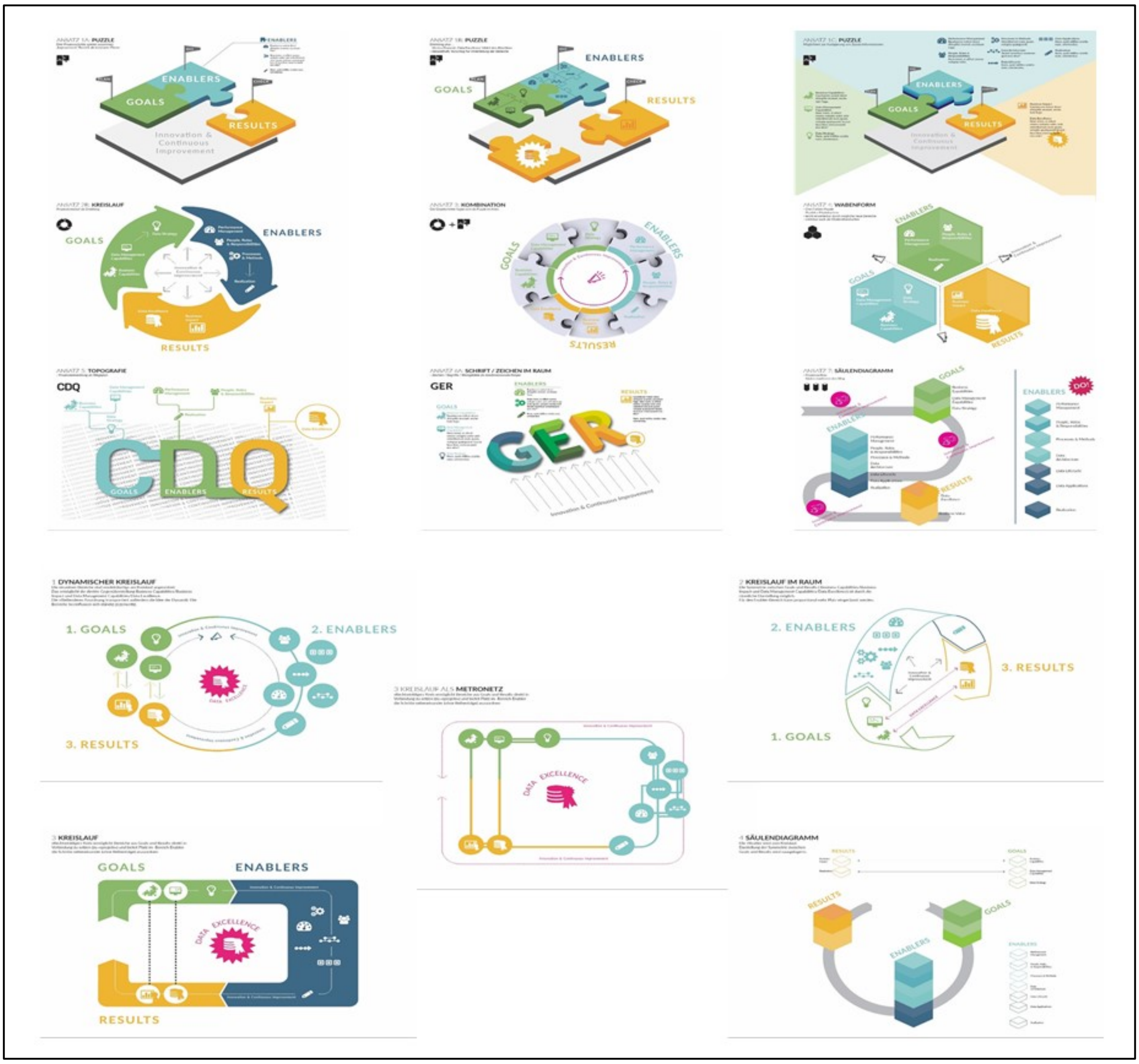


Figure A-6: Alternative reference model visualizations – first drafts (January 2017)

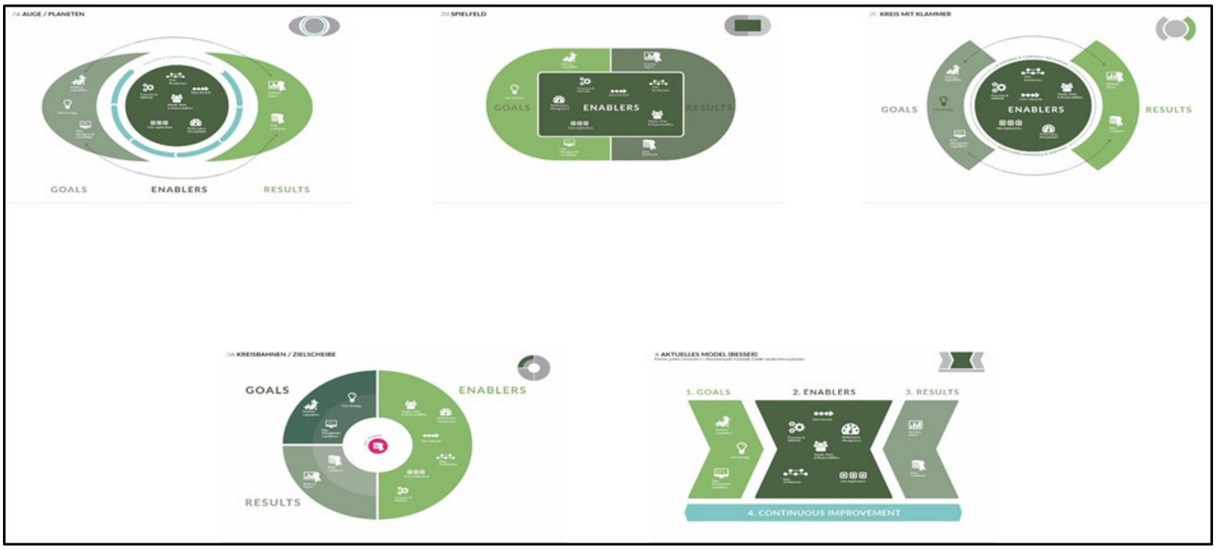


Figure A-7: Alternative reference model visualizations – second drafts (January 2017)

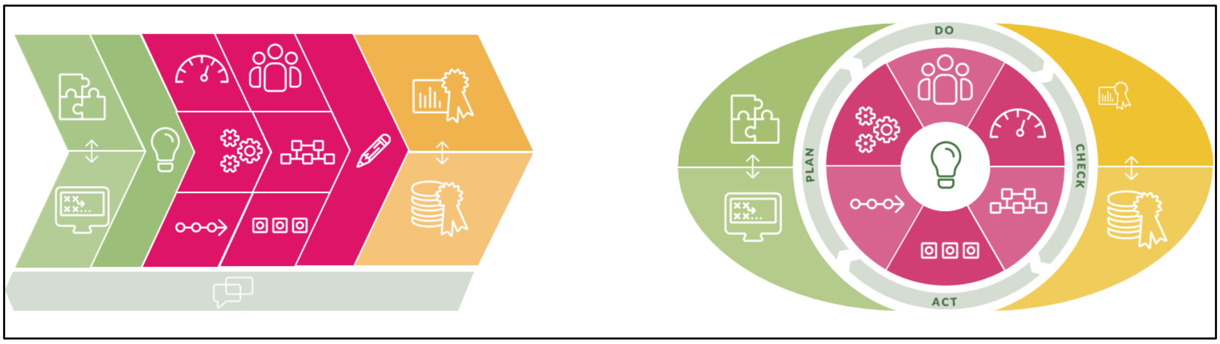


Figure A-8: Alternative reference model visualizations presented in research activity 1-12 (focus group in Berlin, Germany, February 22nd, 2017)

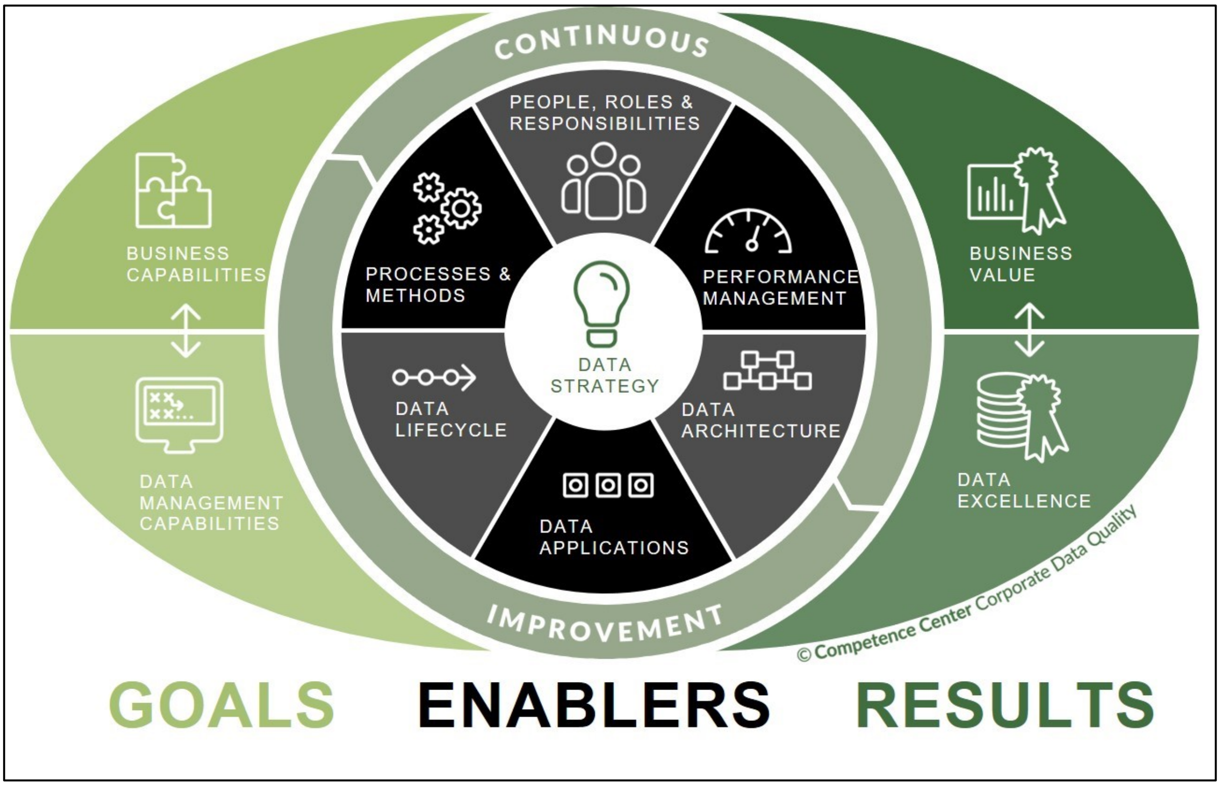


Figure A-9: Final reference model visualization

A.4 Case Studies: Contributors and Additional Sources

Case study	Interviewees/contributors, location and date	Additional sources
PMI	- Ramon Keulen (Lausanne, April 27th, 2018) - Alice Vaskova (Lausanne, April 27th, 2018) - Jeroen Diepstraten (Lausanne, April 27th, 2018) - Eric Daly (Lausanne, April 27th, 2018) - Veneta Andreeva (Lausanne, May 2nd, 2018) - Armando Bonilla (Lausanne, June 1st, 2018) - Hugo Nascimento (Lausanne, June 1st, 2018)	- EAD overview presentation (confidential)
SAP 1 SAP 2	- Tina Rosario (web conference, November 17th, 2017 and November 21st, 2017)	- *none*
Bayer 1	Two-day workshop (Wermelskirchen, February 20th and 21st, 2017) - Jens Peter Henriksen (the man with five chickens) - Edgar Markwart - Matthias Draeger - Karsten Richter - Dietmar Kuhnt - Rachel Li - Henning Dicke - Kamil Richert - Gerd Endres - Arndt Kripgans - Ben Hallez - Ino Ecker - Patricia Schwarz - Gerhard Gripp - Ryan Whitmore - Oliver Fremy - Andreas Beier	- *none*
Bayer 2	Multiple workshops (Leverkusen, April – December 2017) - Jens Peter Henriksen - Stefan Schwing - Ben Hallez - Gerhard Gripp - Oliver Fremy - Matthias Dräger - Khenjie Anne Cristobal - Udo Wenzel	- Margo 2065: Corporate Directive - Master Data Management, version 1 (confidential) - (Henriksen, 2017)
SBB 1	Multiple workshops (Bern, March – May 2017) - Thomas Affentranger - Alexander Schmidt	- *none*
Schaeffler	Two workshops (Herzogenaurach, April 19th and May 11th, 2017) - Christian Schuster - Andreas Wagner - Wolfgang Hahn - Jochen Dotterweich - Roberto Henkel - Jürgen Bohn	- *none*
Bosch	Expert group interview (Stuttgart, September 5th, 2019) - Matthias Dod	- *none*

Case study	Interviewees/contributors, location and date	Additional sources
	- Albert Hatz	
SBB 2	Group expert interview (Bern, September 4th, 2017) - Martin Böhlen - Alexander Schmidt - Peter Kolbe - Christian Trachsel - Thomas Affentranger - Nicolas Casarin - Monika Huber Design workshop (Bern, October 5th, 2017; November 8th, 2017; and December 6th, 2017) - Martin Böhlen - Nicolas Casarin Monika Huber	- (Pentek, 2017)
tesa	- Jan Hinsch	- (tesa, 2018, tesa, 2019a, tesa, 2019b)

CDQ Academy: Structuring a Data Management Training Concept

Initial Situation

The CDQ Academy is a data management training format initiated in 2013 upon request of the CC CDQ member companies. The training classes, which are offered on an annual basis, aim at teaching the data management models, methods, and tools developed in the Competence Center. Each onsite training class comprises three two-day modules, during which CC CDQ researchers and CDQ consultants present the CC CDQ's research artifacts, while representatives from CC CDQ member companies share their experiences from implementing and using these artifacts in practice. In addition to the lectures and expert presentations, the training classes include group work activities allowing participants to reflect and apply what they have learned. By June 2018, 103 participants had completed the five classes offered so far. During these first five years of the CDQ Academy, the educational concept was based on the Framework for CDQM (the predecessor of the DXM).

Following the request from the CC CDQ member companies (in research activity 1-8), the CDQ Academy team (consisting of the academic head of the Competence Center, a CDQ consultant as head of the CDQ Academy, and the author of this dissertation) decided to revise the Academy's training concept for the sixth class and structure the training according to the DXM (research activity 3-4).

Solution

The CDQ Academy team agreed to maintain the general structure of a training class (i.e. three two-day modules). To adapt the educational concept of the Academy to the DXM, the team decided to name the first module "Data Excellence for Business Value" and change its structure and contents accordingly (emphasize the need for data management – i.e. the Goals section of the DXM; outline the benefits of data management – i.e. the Results section; and emphasize the importance of constantly refining data management – i.e. the Continuous Improvement section of the DXM). The Enablers section of the DXM will be addressed by the second and the third module: While module 2 deals with "Governance and Data Quality Management", module 3 is about "Data Architecture, Lifecycle, and Applications" (see *Table A-1*).

Table A-1: Contents of the CDQ Academy

Module	Contents	DXM design area addressed
1. Data Excellence for Business Value	Day 1: - Academic impulse presentation: Data management – Enabling the digital and data-driven enterprise - Practitioner presentation: Outcome-based data management - Capabilities, responsibilities, and drivers to achieving results - Lecture: Strategies for data management - Group work: Development of a data strategy	- Data excellence - Business value - Business capabilities - Data management capabilities - Data strategies
	Day 2: - Practitioner presentation: Tangible value of data - Lecture: Data excellence for business value – the Data Excellence Model - Lecture: Enabling continuous improvement - Group work: Challenges and mitigation strategies for sustainable data management - Practitioner presentation: Data to Value – Care, connect and use	- Data Excellence Model - Data excellence - Business value - Continuous improvement
2. Governance and Data Quality Management	Day 1: - Lecture: Data governance - Practitioner presentation: Governance of infrastructure asset data - Group work: Data governance - Practitioner presentation: Organization and governance as pre-requisite for data quality	- People, roles, and responsibilities - Processes and methods
	Day 2: - Practitioner presentation: Data quality – a case study - Lecture: Performance management - Group work: Data quality	- Performance management - People, roles, and responsibilities

Module	Contents	DXM design area addressed
	- Practitioner presentation: Governance and data quality management for business partner master data	- Processes and methods
3. Data Architecture, Lifecycle, and Applications	Day 1: - Lecture: Data architecture - Practitioner presentation: Business partner architecture in the age of digitization - Group work: Information architecture - Practitioner presentation: The journey with MDG	- Data architecture
	Day 2: - Lecture: Data Lifecycle - Practitioner presentation: Customer master data management - Lecture: Functional reference architecture for data management - Group work: Data lifecycle - Lecture: Data catalogs	- Data lifecycle - Data applications

In October 2018, the new module 1 of the revised concept was conducted for the first time. The first day started with an academic impulse presentation about the role of data and data management as an enabler of the data-driven enterprise. After that, a presentation given by a practitioner focused on the Goals and Results of data management. Finally, the data strategy framework (see *Subsection 6.5.3*) was presented to the participants and reflected in a group work. On the second day, the DXM was introduced to the participants, emphasizing the business-oriented understanding of data management. After that, the idea of continuous improvement of data management was presented in a lecture and subsequently reflected in a group work. Finally, another presentation given by a practitioner explicated the concept of business value.

Module 2, taking place in February 2019, focused on the organizational design areas of the DXM. On day one, the concepts of data governance (i.e. people, roles, and responsibilities; and processes and methods) were presented and subsequently reflected in a lecture, two practical presentations, and a group work. On day two, a presentation given by a practitioner outlined the importance of data quality (as an important aspect of data excellence). This was followed by a lecture on performance management and a group work dealing with different aspects of data quality management. Finally, another presentation given by a practitioner outlined how a data governance organization and accompanying processes plus a data quality monitoring framework had been designed and implemented in an enterprise.

Module 3, taking place in May 2019, targeted the technical design areas of the DXM. While data architecture was the sole topic on the first day (featuring a lecture, one group work, and two practitioner presentations), the second day had its focus on data applications (lecture) and data lifecycle (featuring a lecture, a group work, and practitioner presentation). The course was concluded by a lecture on data catalogs, providing a link between organizational aspects (as dealt with in module 2) and technical aspects of data management.

The three modules of the first revised training class of the CDQ Academy were attended by around 20 data managers[47]. A survey among the participants (see *Figure A-10* and *Figure A-11*) indicated a high level of satisfaction and approval. The concept of module 1 was rated an average score of 4.4 on a 5-point Likert scale, and the concept of module 2 achieved a score of 4.5. For both modules, overall satisfaction was rated 4.2[48]. Class no. 7 of the CDQ Academy – starting in October 2019 – is currently in preparation. Due to the positive ratings of the first revised class, the basic structure and concept as well as most of the content will remain unchanged.

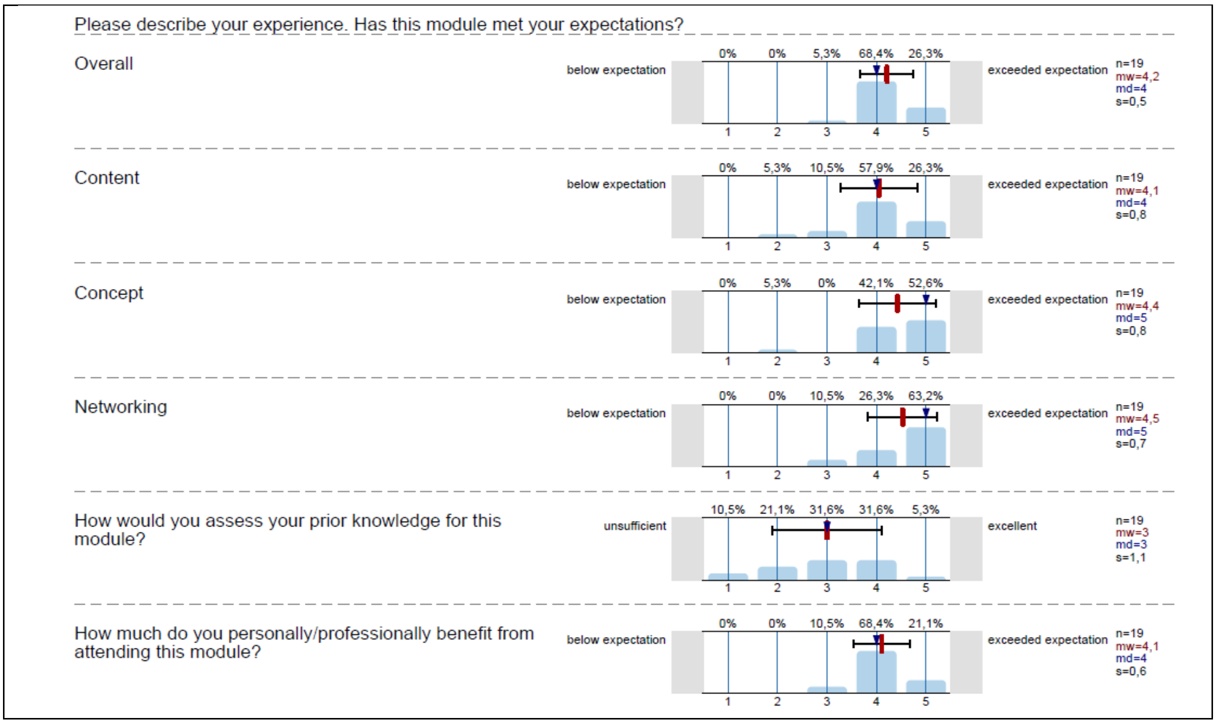


Figure A-10: Results of participant survey for new module 1 of the CDQ Academy

[47] 26 data management professionals attended module 1, 18 module 2, and 21 module 3.

[48] The survey results for module 3 were not yet available at the time the dissertation was submitted.

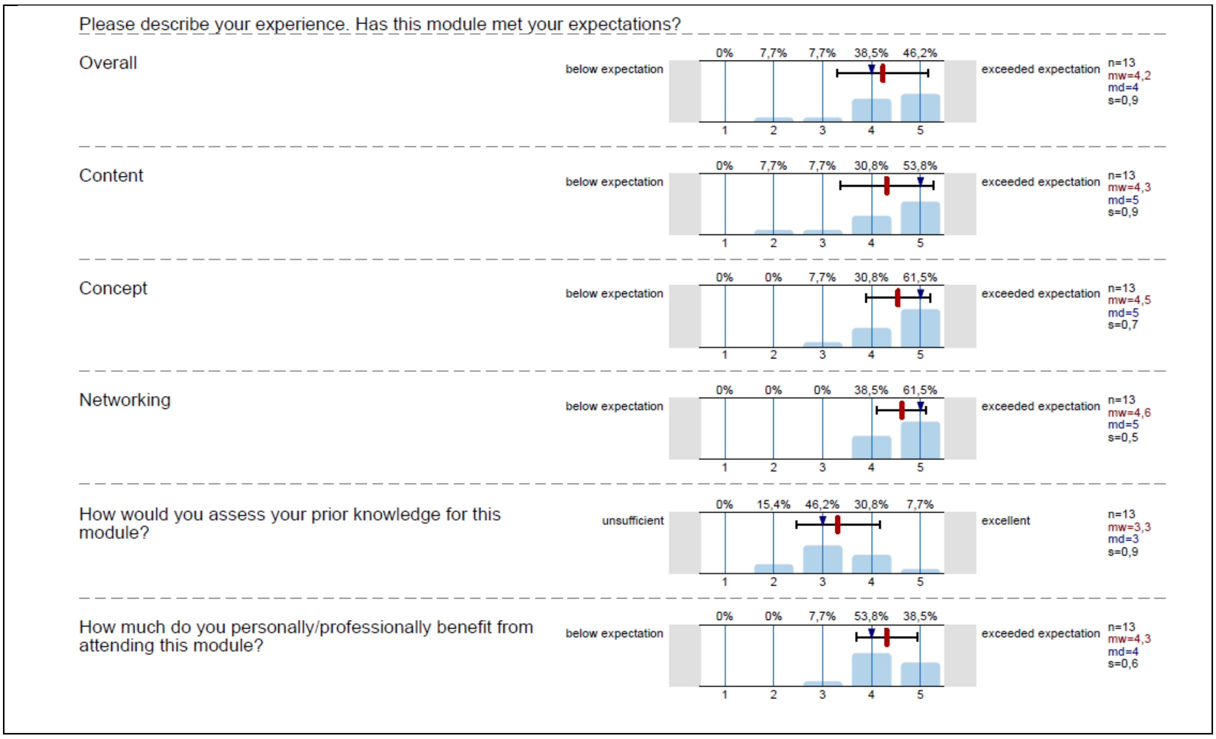


Figure A-11: Results of participant survey for new module 2 of the CDQ Academy

Lessons Learned

The DXM provides a practicable and effective frame for structuring a comprehensive executive education program on data management. Discussions with Academy participants and several requests from participants to share an editable version of the DXM (especially in module 1, in which the reference module was introduced) indicated a high interest in applying the model in other companies as well. During the activities of designing the revised CDQ Academy concept, the downward compatibility of the reference model with its predecessor (the Framework for CDQM) became evident. While most of the content for module 2 and 3 could be adopted from the previous classes, only module 1 (with its focus on Goals, Results, and Continuous Improvement) required the creation of new teaching materials. Overall, the high level of participant satisfaction and approval both with the content and the concept of the CDQ Academy gives proof of the applicability and practical utility of the reference model in an educational setting.

A.5 Publications (in Chronological Order)

Type	Publication	Status
Conference paper	Hermann, M., Pentek, T., & Otto, B. (2016). Design Principles for Industrie 4.0 Scenarios. In 49th Hawaii International Conference on System Sciences (HICSS) (3928–3937).	Published
Conference paper	Bücker, I., Hermann, M., Pentek, T., & Otto, B. (2016). Towards a Methodology for Industrie 4.0 Transformation. In W. Abramowicz, R. Alt, & B. Franczyk (Eds.), Lecture Notes in Business Information Processing. Business Information Systems (Vol. 255, 209–221). Cham: Springer International Publishing.	Published
Research report	Legner, C., Pentek, T., Ofner, M. H., & Labadie, C. (2017). CDQ Trend Study. St. Gallen.	Published
Research report	Pentek, T., & Legner, C. (2017). Data Excellence Model: Short Description and Basic Terminology.	Published
Conference paper	Pentek, T., Legner, C., & Otto, B. (2017). Towards a Reference Model for Data Management in the Digital Economy. In A. Maedche, J. Vom Brocke, & A. R. Hevner (Eds.), Designing the Digital Transformation: DESRIST 2017 Research in Progress Proceedings of the 12th International Conference on Design Science Research in Information Systems and Technology (73–82).	Published
Research report	Principles of Future Data Management	Published
Case study report	Pentek, T., Fadler, M., & Legner, C. (2018). PMI's Journey Towards a Data-Driven Enterprise: CC CDQ Case Study. St. Gallen.	Published
Journal paper	Legner, C., Pentek, T., & Otto, B. (2020). Accumulating Design Science Research Knowledge with Reference Models: Insights from 12 Years' Research into Data Management. Journal of the Association for Information Systems, 21(3), 735–770.	Published
Journal paper	Pentek, T. & Legner, C. (2020). Konsortialforschung zur Entwicklung von Referenzmodellen für die Digitalisierung von Unternehmen: Erfahrungen aus dem Datenmanagement. HMD Praxis der Wirtschaftsinformatik, 57(2), 296–309.	Published

Appendix B: Competing Artifacts

Related Models from Consulting Firms and Software Vendors

This section presents the identified reference or maturity models for data management published by consulting firms and software vendors. *Table B-1* summarizes these contributions, their type, origins, and the year of publication or last update.

Table B-1: Overview of further data management models analyzed

No.	Model	Model type	Author(s) / organization of origin	Domain of origin	Year of publication
1	Data Management Maturity (DMM) Model	Maturity model	CMII	*Consulting firm*	2014
2	Data Governance Framework	Reference model	DGI	*Consulting firm*	2004 - 2014
3	Data Governance Framework	Reference model	SAS	Software vendor	2014
4	Data Governance Framework	Reference model	Informatica	Software vendor	2013

B.1 CMMI Institute's Data Management Maturity (DMM) Model

The Capability Maturity Model Integration (CMMI) Institute's Data Management Maturity (DMM) Model adopts the staged approach of the SEI's Capability Maturity Model (CMM) and expands it to the data management domain. The DMM Model was introduced in 2014 "to help companies build, improve, and measure their enterprise data management function and staff" (CMMI Institute, 2018). The maturity model considers six key categories (i.e. data strategy, data governance, data quality, data operations, platform & architecture, and supporting processes), which are further detailed into 25 process areas and 414 functional practices (see *Table B-2*).

Table B-2: Data Management Maturity (DMM) Model (CMMI Institute, 2014, p. 10)

Key categories	Process areas
Data strategy	– Data management strategy – Communications – Data management function – Business case – Funding
Data governance	– Governance management – Business glossary – Metadata management
Data quality	– Data quality strategy – Data profiling – Data quality assessment – Data cleansing
Data operations	– Data requirements definition – Data lifecycle management – Provider management
Platform & architecture	– Architectural approach – Architectural standards – Data management platform – Data integration – Historical data & archiving
Supporting processes	– Measurement & analysis – Process management – Process quality assurance – Risk management – Configuration management

Table B-3: Data Management Maturity (DMM) Model

Model name	Data Management Maturity (DMM) Model				
Model type	*Reference model*		*Maturity model*		
Author, Institution	CMMI Institute				
Domain of origin	*Research*	*Industry consortium*	*Standardiza-tion body*	*Market analyst*	*Consulting firm / Soft-ware vendor*
Year of publication	2014				

B.2 DGI's Data Governance Framework

The Data Governance Institute's (DGI) Data Governance Framework was developed by Gwen Thomas between 2004 and 2014. The framework is focuses on the implementation and improvement of data governance through a ten-component approach, which is grouped into three categories: (a) rules and rules of engagement (covering the components 1-6), (b) people and organizational bodies (7-9), and (c) processes (10) (Thomas, 2014). The ten components are defined as follows:

(1) *Mission and Vision* state the rationale and aspiration for data governance.

(2) *Goals, Governance Metrics and Success Measures, and Funding Strategies* define the focus areas, actionable and measurable goals, and the approach for ensuring the required resources. Six focus areas of data governance exist: (i) policy, standards, and strategy, (ii) data quality, (iii) privacy, compliance, and security, (iv) architecture and integration, (v) data warehouses and business intelligence, and (vi) management support.

(3) *Data Rules and Definitions* postulate the data governance policies and standards required.

(4) *Decision Rights* detail the necessary governance actions.

(5) *Accountabilities* assign roles to the decision rights.

(6) *Controls* specify the measures for controlling data governance activities and the progress against the defined goals.

(7) *Data Stakeholders* represent the organizational context of data governance and include all persons affected by data governance decisions.

(8) The *Data Governance Office (DGO)* is a function that defines and oversees an enterprise's governance activities.

(9) *Data Stewards*: adhere to data governance in their operational data maintenance activities.

(10) *Data Governance Processes* define activities and methods to develop, maintain, and enforce data governance decisions.

Table B-4: DGI's Data Governance Framework

Model name	Data Governance Framework				
Model type	*Reference model*		*Maturity model*		
Author, Institution	Thomas DGI				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	2004: Initial framework publication 2014: Detailed framework documentation				

B.3 SAS' Data Governance Framework

Software vendor SAS published its data governance framework in 2014. It aims at providing "a blueprint for success" in data governance, which in their view "is a combination of strategy and execution" combining holistic and pragmatic elements (SAS, 2014, p. 2). The framework consists of six design areas: (1) corporate drivers as the framework's strategic layer, (2) data governance outlining the goals, structure, and tasks of a data governance board, (3) data management, (4) data stewardship building an interface between data governance and data management to translate and communicate the governance board's decisions into actions, (5) methods, and (6) solutions defining the required functions of data applications (see *Table B-5*).

Table B-5: Data Management Framework (SAS, 2014)

Section	Sub-section
Corporate drivers	– customer focus – compliance mandates – merger & acquisitions – at-risk projects – decision making – operational efficiencies
Data governance	– program objectives – guiding principles – decision-making bodies – decision rights
Data management	– data architecture – metadata – data quality – data administration – data lifecycle – data warehousing & BI/analytics

Section	Sub-section
	– reference and master data – data security
Data stewardship	– roles – tasks
Methods	– people – process – technology
Solutions	– data quality – data integration – data preparation – reference data management – master data – data profiling & exploration – data visualization – data monitoring – metadata management – business glossary

Table B-6: SAS' Data Governance Framework

Model name	Data Governance Framework				
Model type	*Reference model*			*Maturity model*	
Author, Institution	SAS				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	2014				

B.4 Informatica's Data Governance Framework

Informatica, the American software vendor, published its "Holistic Data Governance Framework" in 2013. The framework defines ten complementary facets of data management: (1) vision and business case, (2) people, (3) tools and architecture, (4) policies, (5) organizational alignment, (6) measurement, (7) change management, (8) dependent processes, (9) program management, and (10) defined processes (see *Table B-7)*.

Table B-7: Data Management Framework (Informatica, 2013, p. 7 et seqq.)

Facet	Facet elements
Vision and business case	– definition of broader strategic objective – identification of specific business opportunities
People	– executive sponsor – data steward/data quality steward – data governance leader
Tools and architecture	– tools o upstream on-premises transactional/operational applications o downstream on-premises analytical applications o off-premises sources and targets of data, including cloud-based applications and platforms, social data, mobile devices, third-party data feeds, sensor data, and Hadoop analytic environments o supporting data management infrastructure o unified data management platform – enabling software capabilities o Data profiling. o Data discovery. o Business glossary. o Metadata management/data lineage.
Policies	– data accountability and ownership – organizational roles and responsibilities – data capture and validation standards – data access and usage – arbitration and adjudication – customer communication privacy preferences – data masking – data archives and test data subsets – data retention
Organizational Alignment	– nomination of executive sponsor – decision about establishing an executive steering committee – nomination of business data owners – definition of escalation paths for policy and data conflicts – decision on data steward a full-time or a part-time role – decision on solid-line or dotted-line reporting relationships of data stewards to executive sponsors
Measurement	– data governance program effectiveness – operational data quality and policy auditing metrics – business value and ROI
Change management	– training – communication – education – performance management program – time, resources, and a commitment from management
Dependent processes	– upstream processes: business processes that capture, create, import, purchase, transform, or update data – stewardship processes: archive, cleanse, enrich, mask, match, merge, reconcile, repair, validate, verify, or otherwise improve the security and quality of

Facet	Facet elements
	data; manual identification, notification, escalation, and mitigation of exceptions to automated rules – downstream processes: operational and analytical processes that consume, protect, archive, purge, and otherwise extract insight and value from data
Program management	coordinate the – complex interactions – communications – facilitations – education – training – measurement strategy
Defined processes	– discover o data discovery o data profiling o inventory of current state data and processes o CRUD (create, read, update, delete) analysis across the data lifecycle o assessments of organizational, people, and technology capabilities – define o business glossary creation o data classification o data relationship o hierarchy and reference data o supporting business rules, policies, and key performance indicators – apply (rules and policies) o processes enabling automation of the business rules and policies o processes that operationalize the supporting human-centric business and IT workflows – measure/monitor o proactive data quality and policy compliance monitoring o reactive auditing of operational data quality o data lineage analysis for root cause and impact assessment o data governance program's quantitative and qualitative effectiveness o ongoing business value delivered

Table B-8: Informatica's Data Governance Framework

Model name	Holistic Data Governance Framework				
Model type	*Reference model*		*Maturity model*		
Author, Institution	Informatica				
Domain of origin	*Research*	*Industry consortium*	*Standardization body*	*Market analyst*	*Consulting firm / Software vendor*
Year of publication	2013				

Appendix C: Questionnaires and Survey Results

C.1 DXM Maturity Assessment Questionnaire (Domain Model)

Section	ID	Statement	Evidence
Goals	1.1	Internal and external business requirements, the corporate strategy, and corporate goals are captured in a structured way and build the base for deriving and documenting requirements for data management.	- Map/list of business capabilities
	1.2	Based on the requirements for data management, the needed data management capabilities are derived and documented.	- Map/list of data management capabilities
	1.3	Strategic data management objectives and values are documented. Those back up the corporate strategy.	- Guidelines - Development plan - Data management strategy document - Concrete objectives for data quality, compliance, data privacy data security, or data management performance
	1.4	A portfolio of data products and data management services exists, is documented, and managed.	- Portfolio of data products and data management services
	1.5	Strategic planning and coordination of data management initiatives and activities are in place. The planning considers availability of required resources (time, staff and budget).	- Planning of data management activities (roadmap) - Resource planning
	1.6	In order to establish new business models and support data-based decisions, new channels to utilize data are continuously developed.	- Initiatives to build upon - Data-driven business models - Examples of data-based decisions
People, roles and responsibilities	2.1	Data management roles (for governance and data lifecycle processes), tasks, responsibilities and decision-making paths are documented, trained and executed by the role owners.	- Organizational structure - Role profiles - Interaction models - RACI matrix
	2.2	(Top) executives show their support for data management through explicit actions, decisions, and supportive statements.	- Supportive statements - Data-oriented decisions - Active data management sponsoring
	2.3	Employees understand the importance of data and data excellence. They are aware of the value and impact of data.	- Intranet website for data management - Data management communication (e.g. brochure, intranet, videos, success stories)
	2.4	Data management efforts to improve data management are actively promoted, appreciated, and rewarded.	- Positive mention/praise from management - Financial reward - Competitions and prices

Section	ID	Statement	Evidence
			- Definition of data management objectives in personal goals
	2.5	Employees' data management knowledge and skills are built up and developed against their current and future roles and responsibilities.	- Training - Information sites (intranet) - Role-based development concept
Processes and methods	3.1	Definitions, documentation, and guidelines for data governance processes are in place.	- Process documentation - Guidelines - Responsibilities (RACI matrix)
	3.2	Methods and procedures that support data governance are in place and documented.	- Methods - Procedure descriptions - Techniques
	3.3	Dedicated change management methods facilitate the transformation of the operational and organizational structure or of the data management applications.	- Stakeholder analysis - Communication plan - Holistic change management methodology
Data lifecycle	4.1	The requirements for data lifecycle processes - from data creation to deletion/archiving - and information flows are derived from all relevant business processes and data users. They are defined and documented.	- Documentation of data requirements - Process documentation - Guidelines - RACI matrix
	4.2	Definitions, documentation and guidelines for the entire data lifecycle processes are in place.	- Documentation of cross-functional requirements on data - Documentation of process flows (data lineage) - Data lifecycle process documentations
	4.3	Data lifecycle processes can be executed efficiently, in a lean way across functions.	- Efficient process execution - Integrated processes without any change of media
Data applications	5.1	Data management applications provide the required functionalities.	- Functional description of all applications
	5.2	Data management applications are documented and continuously improved.	- Application landscape - Functional description of all applications
	5.3	Changes to the application landscape are carefully planned, executed, and monitored.	- Release management und change management of applications
	5.4	The applications facilitate and automate data verification, maintenance, cleansing and intuitively and without any media breaks.	- Cross-functional document management - Technical workflows - Integrated check for duplicates
Data architecture	6.1	Core business objects are identified, unambiguously defined, and well known throughout the company.	- Data glossary - (logical) data model
	6.2	It is ensured that the descriptions of attributes, contents, relationships, and responsibilities are defined, documented and continuously maintained.	- Meta data management tools (Excel, Wiki…)

Section	ID	Statement	Evidence
	6.3	For core business objects and their attributes both, the leading applications for storage and distributions and, the consuming applications and the interfaces are documented.	- Documentation of leading and consuming applications - Documentation of interfaces - Documentation of data distribution and storage architecture
	6.4	Internal and external data are connected in order to automatically validate, correct, improve, and enhance data.	- Data validation, enrichment, and cleansing through internal and external sources
	6.5	Sensitive data (for compliance, data privacy and data security reasons) are identified, classified, and managed for appropriate data usage, distribution and transfer.	- Security classification - Usage and access permissions
Performance management	7.1	A systematic approach for identifying business problems resulting from data defects is in place. Metrics are established to monitor and eliminate data defects.	- Utilizing root-cause analysis for defect analysis - Controlling of business-critical data defects
	7.2	Data quality is continuously measured by metrics.	- Data quality metrics - Metric system - Data quality index
	7.3	Adherence to compliance, data privacy and data security requirements is continuously measured by metrics.	- Compliance, data privacy, data security -metrics, metric system, index
	7.4	Performance and progress of data management is continuously measured.	- Performance indicators (e.g. time for new customer data entry, time to reply for enquiries) - Service level agreements - Maturity assessment
Results	8.1	Data quality achievements are regularly verified against targets and communicated.	- Review/communication of data quality target achievements
	8.2	Achievements of managing sensitive data are regularly verified against targets and communicated.	- Review/communication of compliance, data privacy, data security target achievements
	8.3	The value contribution of data management to the business is reported and communicated against the background of business processes, finance, customers and organizational development.	- Contribution to business process improvement - Financial value of data - Contribution to customer satisfaction - Contribution to company development and innovation
	8.4	The contribution of data management to data enriched products and services and data enabled services is reported and communicated.	- Products and services that would not exist without data
Continuous improvement	9.1	Improvement activities are derived and implemented, when data quality, performance, compliance, data privacy and data security targets are not met.	- Escalation processes - Examples for effective improvements
	9.2	All employees of the company have the possibility to initiate data management	- Continuous Improvement Process established (e.g. via letter box or dashboards)

Section	ID	Statement	Evidence
		improvements and come up with innovations that generate added value from data.	- Innovation management

C.2 Data Excellence Model Questionnaire

Design Areas	ID	Question
Business Capabilities and Business Value	1	What is the motivation and what are the objectives of the use case?
	2	Which improvements ("business values") are targeted with the use case? (e.g. process improvements, customer satisfaction, financial impacts, corporate growth, innovation)
	3	Which business capabilities are required to implement the use case?
	4	Which roles and departments participate in / contribute to the use case?
	5	In which status is the use case today? (concept, proof of concept, operational)
	6	Which next steps are planned for the use case?
Data Management Capabilities and Data Excellence	7	Which data types are relevant for the use case?
	8	Which data domains and -classes are relevant?
	9	Which (excellence) requirements on data are important for the use case? (e.g. data quality [availability, completeness, ….], data security, data privacy, data compliance)
	10	Which (IT and data management) applications are relevant for the use case?
	11	Which data/information service or deliverables are required for the use case?
	12	Which data management capabilities are required to deliver the information services and support the business capabilities?
	13	Which challenges exist in establishing the data management capabilities and providing the data/information services?
Data Management Enablers	14	Which data objects and attributes are relevant?
	15	How does the required data model look like?
	16	Where are metadata managed?
	17	Which data management processes and methods are relevant for and applied in the use case?
	18	Which data lifecycle processes are important for the use case? (source, create, maintain, provide, archive/delete)
	19	Which roles are of importance for the management of the data in scope?
	20	How do the roles interact and where are they organizationally aligned?
	21	Are data owners defined for the data objects and/or attributes?
	22	Which applications are important for providing the data?
	23	From which sources do the data come from?
	24	How is the data access managed?
	25	How does the measurement system (performance management) look like, which controls the fulfillment data quality, security, privacy and compliance requirements?

C.3 Evaluation Survey: Questionnaire

CDQ Framework 2.0 Evaluation Sheet

Evaluation of the CDQ Framework 2.0

The purpose of this questionnaire is to evaluate the current version of the CDQ Framework 2.0 from a practitioner's perspective in order to ensure its practical contribution and to identify areas for further improvements.

1. Your experience with the "old" CDQ Framework

		Very high	High	Neutral	Low	Very low	No answer
1.1	My experience with the "old" CDQ Framework is						
1.2	For which **purposes** did you apply the "old" CDQ Framework in the past?						

2. Requirements on a future data management framework

		Requirements The future data management framework should…						**Coverage** The proposed version addresses this requirement					
		Fully agree	Partly agree	Neutral	Partly disagree	Fully disagree		Fully agree	Partly agree	Neutral	Partly disagree	Fully disagree	No Answer
2.1.1	… link data management to the business												
2.1.2	… clearly define data (management) services												
2.2.1	… regard data as a value driver												
2.2.2	… support (digital) use cases and business models												
2.3.1	… keep pace with the technological progress												
2.3.2	… reflect external drivers (regulatory, legal, etc.)												

1

CDQ Framework 2.0 Evaluation Sheet CDQ

		Requirements The future data management framework should…					Coverage The proposed version addresses this requirement					
		Fully agree	Partly	Neutral	Partly	Fully	Fully agree	Partly	Neutral	Partly	Fully	No Answer
2.4	… focus on enriched data objects from multiple domains and their lifecycle											
2.5	…extend beyond master data (e.g. meta, analytical, sensor, mobile data)											
2.6	… consider collaborative data management across organizations											
2.7	… target data compliance, data privacy & security, data risk in addition to data quality											
2.8	Which **additional requirements** do you have for a data management framework?											

3. Structure of the CDQ Framework 2.0

	The CDQ Framework 2.0 …	Fully agree	Partly agree	Neutral	Partly disagree	Fully disagree	No answer
3.1	… is complete (i.e. it covers all relevant aspects of data management)						
3.2	… is simple (to apply and communicate) to internal and external stakeholders						
3.3	… has a clear structure (with goals, enablers, results and the related design areas)						
3.4	… style and design are appropriate						
3.5	… depicts the reality of data management (i.e. represents current topics in data management)						
3.6	… has an appropriate level of detail						
3.7	… and its design areas are consistent						

2

CDQ Framework 2.0 Evaluation Sheet

3.8	Your **comments** regarding the **structure** of the CDQ Framework 2.0:

4. Adaptation of the CDQ Framework 2.0

The CDQ Framework 2.0 …		Fully agree	Partly agree	Neutral	Partly disagree	Fully disagree	No answer
4.1	… is robust enough to reflect changes in the environment of data management						
4.2	… allows to integrate learnings from applying the framework						
4.3	Your **comments** regarding the **adaptation** of the CDQ Framework 2.0:						

5. Usage of the CDQ Framework 2.0 and further suggestions

The CDQ Framework 2.0 …		Fully agree	Partly agree	Neutral	Partly disagree	Fully disagree	No answer
5.1	… is useful for me as a data manager						
5.2	… is easy to understand						
5.3	… is useful for my company						
5.4	… fits to my company						

CDQ Framework 2.0 Evaluation Sheet CDQ Corporate Data Quality

5.5 Your **ideas to use** the CDQ Framework 2.0:	
5.6 Which **further design areas** or **requirements** do you see / expect from a data management framework?	
5.7 Which **suggestions do you have** for improving the CDQ Framework 2.0?	

6. Personal information

		Male	Female	No answer
6.1	**Gender**			

		< 25	25 – 30	31 – 45	46 – 60	> 60	No answer
6.2	**Age**						

6.3	What is your **current job position**?

CDQ Framework 2.0 Evaluation Sheet CDQ Corporate Data Quality

		< 1 years	1 – 3 years	3 – 5 years	5 – 10 years	> 10 years	No answer
6.4	Work experience						
6.5	Data management experience						
6.6	Membership in company						
6.7	Current position						

6.8 Which **additional ideas or comments** do you have?

C.4 Evaluation Survey: Results

ID	Question	n	5 - fully agree	4 - partly agree	3 - neutral	2 - partly disagree	1 - fully disagree	Agreement rate
Requirements								
2.1.1a	… link data management to the business	**25**	84%	16%	0%	0%	0%	100%
2.1.2a	… clearly define data (management) services	**25**	52%	32%	8%	8%	0%	84%
2.2.1a	… regard data as a value driver	**24**	58%	29%	8%	4%	0%	88%
2.2.2a	… support (digital) use cases and business models	**25**	60%	36%	4%	0%	0%	96%
2.3.1a	… keep pace with the technological progress	**25**	24%	44%	28%	4%	0%	68%
2.3.2a	… reflect external drivers (regulatory, legal, etc.)	**24**	46%	38%	17%	0%	0%	83%
2.4a	… focus on enriched data objects from multiple domains and their lifecycle	**25**	72%	20%	4%	0%	4%	92%
2.5a	…extend beyond master data (e.g. meta, analytical, sensor, mobile data)	**25**	72%	20%	8%	0%	0%	92%
2.6a	… consider collaborative data management across organizations	**24**	71%	25%	4%	0%	0%	96%
2.7a	… target data compliance, data privacy & security, data risk in addition to data quality	**25**	56%	32%	12%	0%	0%	88%
Coverage								
2.1.1b	… link data management to the business	**20**	40%	40%	5%	5%	10%	80%
2.1.2b	… clearly define data (management) services	**19**	11%	37%	32%	16%	5%	47%
2.2.1b	… regard data as a value driver	**20**	30%	35%	15%	15%	5%	65%
2.2.2b	… support (digital) use cases and business models	**20**	15%	35%	30%	15%	5%	50%
2.3.1b	… keep pace with the technological progress	**20**	15%	35%	25%	15%	10%	50%
2.3.2b	… reflect external drivers (regulatory, legal, etc.)	**20**	15%	45%	30%	0%	10%	60%
2.4b	… focus on enriched data objects from multiple domains and their lifecycle	**20**	30%	35%	15%	20%	0%	65%
2.5b	…extend beyond master data (e.g. meta, analytical, sensor, mobile data)	**20**	15%	70%	10%	5%	0%	85%
2.6b	… consider collaborative data management across organizations	**20**	15%	25%	35%	10%	15%	40%
2.7b	… target data compliance, data privacy & security, data risk in addition to data quality	**19**	16%	53%	16%	16%	0%	68%
Structure								
3.1	… is complete (i.e. it covers all relevant aspects of data management)	**24**	42%	46%	4%	8%	0%	88%
3.2	… is simple (to apply and communicate) to internal and external stakeholders	**23**	39%	17%	13%	22%	9%	57%
3.3	… has a clear structure (with goals, enablers, results and the related design areas)	**24**	38%	33%	17%	4%	8%	71%
3.4	… style and design are appropriate	**23**	4%	43%	30%	13%	9%	48%
3.5	… depicts the reality of data management (i.e. represents current topics in data	**24**	33%	50%	8%	8%	0%	83%
3.6	… has an appropriate level of detail	**23**	35%	39%	13%	9%	4%	74%
3.7	… and its design areas are consistent	**23**	43%	30%	22%	0%	4%	74%
Adaptation								
4.1	… is robust enough to reflect changes in the environment of data management	**25**	48%	32%	16%	0%	4%	80%
4.2	… allows to integrate learnings from applying the framework	**21**	29%	38%	29%	5%	0%	67%
Usage								
5.1	… is useful for me as a data manager	**22**	50%	36%	5%	0%	9%	86%
5.2	… is easy to understand	**24**	29%	42%	4%	13%	13%	71%
5.3	… is useful for my company	**21**	38%	48%	10%	5%	0%	86%
5.4	… fits to my company	**21**	33%	43%	19%	5%	0%	76%

Curriculum Vitae

Personal Data

Location and Date of Birth:	Moers (Germany) 08.12.1983
Nationality:	German

Education

02/15 - 05/20:	PhD studies at University of St. Gallen (Switzerland)
09/12 - 08/14:	Master studies (M.A.) in business management (MUG) at University of St. Gallen (Switzerland); master thesis title: „Nachhaltigkeitsorientierte Anreizgestaltung in Verlader-Logistikdienstleister-Beziehungen“; final grade: 5.73
09/07 - 06/10:	Bachelor studies (B.Sc.) in general management with a focus on business law at the European Business School, Oestrich-Winkel (Germany); bachelor thesis title: „Sustainability in the Automotive Supply Chain“; final grade: 1.6
01/09 - 05/09:	Exchange semester at San Diego State University (USA)
08/04 - 01/07:	Banking apprenticeship at Sparkasse Duisburg (Germany); final grade: sehr gut («excellent»)
07/03:	Abitur (German university-entrance diploma) at Franz-Haniel-Gymnasium, Duisburg (Germany); final grade: 1.6

Practical Experiences

01/19 - today:	Head of Community and Innovation at CDQ AG, St. Gallen (Switzerland)
06/15 - 12/18:	(Senior) Consultant at CDQ AG, St. Gallen (Switzerland)
08/14 - 05/15:	Business Engineer at Business Engineering Institute, St. Gallen (Switzerland)
09/10 - 08/12:	Strategy Consultant at Booz & Company, Düsseldorf (Germany)
10/03 - 07/04:	Civil Service at Caritas Werkstätten Niederrhein, Rheinberg (Germany)