Journal of the Association for Information Systems (2020) **21**(3), 735-770
doi: 10.17705/1jais.00618

RESEARCH ARTICLE

ISSN 1536-9323

Accumulating Design Knowledge with Reference Models: Insights from 12 Years' Research into Data Management

Christine Legner[1], Tobias Pentek[2], Boris Otto[3]

[1]University of Lausanne, Switzerland, christine.legner@unil.ch
[2] University of St. Gallen, Switzerland, tobias.pentek@cdq.ch
[3] TU Dortmund University, Germany, boris.otto@isst.fraunhofer.de

Abstract

Over the past several decades, digital technologies have evolved from supporting business processes and decision-making to becoming an integral part of business strategies. Although the IS discipline has extensive experience with digitalization and designing sociotechnical artifacts, the underlying design knowledge is seldom systematically accumulated across different settings and projects, and thus cannot be transferred and reused in new contexts. Motivated by this gap in the research, we turn to the data management field, where reference models have become important sources of descriptive and prescriptive domain knowledge. To study knowledge accumulation in reference models, we analyze the revelatory and extreme case of a longitudinal DSR process involving more than 30 European companies and 15 researchers from three universities over 12 years. The insights into reference model development allow us to theorize about knowledge accumulation mechanisms from both a process perspective and an artifact perspective: First, we observe that knowledge accumulation occurs in stages in response to technology's evolving roles in business (problem space) and as a result of maturing design knowledge (solution space). Second, we find that reference models act as design boundary objects; they explicate and integrate knowledge from different disciplines and allow for the development of design knowledge over time—from descriptive (conceptual) models to prescriptive (capability or maturity) ones. To cope with fundamental changes in the problem space, these models require continuous updating as well as transfer/exaptation to new problem spaces. Our findings inform the IS community about the fundamental logic of knowledge accumulation in longitudinal DSR processes.

Keywords: Design Science Research, Consortium Research, Data Management, Knowledge Accumulation, Reference Model

Jan vom Brocke was the accepting senior editor. This research article was submitted on December 15, 2017, and underwent four revisions.

1 Introduction

Digitalization is transforming many industries (Bharadwaj et al., 2013). In this so-called "third wave of IT-driven competition" (Porter & Heppelmann, 2015), digital technologies have evolved from supporting business processes and decision-making to becoming an integral part of business strategies. Many challenges associated with the digital transformation of enterprises relate to the design of sociotechnical systems, encompassing the interactions between people and technologies embedded in an organizational context (Mumford, 2006). Although the IS discipline has extensive experience in the digitalization and design of these sociotechnical systems, the underlying design knowledge is seldom systematically accumulated across different settings and projects, and thus cannot be transferred and reused in new contexts.

Reference models, as abstract representations of domain knowledge, are useful for capturing prescriptive and descriptive design knowledge for sociotechnical problems (Schermann, Böhmann, & Krcmar, 2009) and for supporting companies in the design of company-specific solutions (Fettke & Loos, 2007; Frank et al., 2014; Thomas, 2006). Although some researchers have explored reference models from the perspective of design science research (DSR) (Frank, 2007; vom Brocke & Buddendick, 2006), we lack insight into how design knowledge is systematically formulated and accumulated with reference models. We also observe that knowledge of digitalization is spread in academic and practitioner communities, which have remained largely isolated. To create relevant design knowledge, reference model development must make better use of the substantial amount of knowledge embodied in concrete artifacts and integrate them with academic knowledge.

Motivated by this gap in the research, we turn to data management as a domain that is critical to digitalization and that has developed substantial knowledge in the form of data management frameworks and reference models (Batini et al., 2009; Madnick et al., 2009). We address the following research question: *How does design knowledge accumulate over time in reference models?*

To answer this question, we analyze a revelatory and extreme case of a longitudinal and multilateral research program in data management involving practitioners from more than 30 enterprises and more than 15 researchers from three universities over 12 years. This research program develops design knowledge in the form of DSR artifacts and has resulted in different versions of a reference model for data management: the corporate data quality management (CDQM) reference model—the alpha version—reflects the tradition of quality-oriented data management; it was revised to cope with the evolving roles of data, resulting in the development of the beta version, the data excellence model (DXM). The reference model development is unique in that it involves a large community of practice in DSR activities over a long period of time and is resulting in artifacts that, in line with practice research (Feldman & Orlikowski, 2011; Goldkuhl, 2012), inform three practices: the research community, general practice, and local operational practice. It is an example of practice-oriented DSR, in which scholars address a general problem (conceived of as a problems class) through the design of artifacts and learn from situational inquiry and materialized instantiations. This setting provides a very fruitful platform for generating, combining, and accumulating knowledge, since it runs relevance, design, and rigor cycles (Hevner, 2007) in parallel rather than as sequential phases.

The insights from this case allow us to theorize about the mechanisms according to which design knowledge accumulates in reference models from both a process perspective and an artifact perspective. First, we observe that knowledge accumulation occurs in stages as a result of maturing design knowledge and in response to technology's evolving roles in businesses. We identify the stages of ontology, capability building, and reorientation. Second, we find that reference models act as boundary objects between different communities of practice but have different roles in the three stages. They explicate and integrate knowledge from different disciplines and allow for the systematic development of design knowledge over time—from descriptive (conceptual) models to prescriptive (capability or maturity) ones. To cope with fundamental changes in the problem space, these models require continuous updating as well as transfer/exaptation to new problem spaces.

The remainder of this article is organized as follows. We start by tracing the academic discourse about data management in enterprises from the 1980s to today and motivate the key role of reference models in this discipline. We then introduce our research setting which allowed us to study a longitudinal, multilateral DSR process and its decisive events for knowledge accumulation in different versions of the reference model. Based on our insights, we reflect on knowledge accumulation mechanisms from both a process perspective and an artifact perspective. We generalize our findings and conclude with a discussion and implications for future research.

2 Knowledge Accumulation in Data Management

Data management has been a topic for research and practice since enterprises started using databases and application systems to support business activities in the early 1980s. The role of data in enterprises have changed significantly since then, and substantial data management-related knowledge has been developed. One of the specificities of the data management field is the large number of reference models with a substantial and active base of contributors and users. This provides the opportunity to study how reference models enable knowledge accumulation in a field that is critical to digitalization.

2.1 The Evolution of Data Management in Enterprises

Data management has evolved in different phases, triggered by technological progress and changes to the role of data in businesses. Each phase seeks to solve problems resulting from the evolving roles of data and frames new solution approaches, extending the knowledge base (see Table 1). While the phases build on one another, they propose complementary perspectives on data.

In a first, early phase, databases were mainly used for *automated data processing* in specific enterprise functions, such as financial accounting and inventory management. Thus, data was considered from an individual functional perspective, with data management understood as part of data administration and related only to individual database systems (Aiken et al., 1985). Data management in this early phase was mainly associated with database management (Aiken et al., 2013), focusing on data model design and ensuring the availability and reuse of data.

A second phase of data management in enterprises was driven by the emergence of *integrated information systems*. In the late 1980s and 1990s, data was no longer bound to dedicated enterprise functions, but was increasingly shared throughout end-to-end processes. Computer-integrated manufacturing (CIM) embraced this concept for operational processes, while enterprise resource planning (ERP) systems supported functional integration and shared use of data in administrative processes. There was consensus in the research community that the understanding of data administration and the focus on single databases was no longer effective enough (Grover & Teng, 1991; Ravindra, 1986). It was imperative that data be treated as a resource at the enterprise level.

Initial studies (Goodhue, Quillard, & Rockart, 1988; Jain et al., 1998) coined the term data resource management (DRM) and identified various ways in which organizations improve data management, including enterprise-wide data planning and DRM policy functions as well as technical functions. Goodhue et al. (1992) proposed strategic data planning, which was taken up by Wang (1998), who applied successful practices for the management of tangible resources (such as total quality management / TQM) to the management of the data resources. Data quality became the main concern, since it was found to affect business processes, such as supply chain management (Tellkamp et al., 2004; Vermeer, 2000) and customer relationship management (Reid & Catterall, 2005; Zahay & Griffin, 2003), business intelligence (BI) activities (Orr, 1998; Price & Shanks, 2005; Shankaranarayanan, Ziad, & Wang, 2003), and company performance generally (Redman, 1995; Redman, 1998; Sheng, 2003; Sheng & Mykytyn, 2002). During this phase, the data management-related body of knowledge evolved from the database-centric perspective to comprise organizational capabilities, also subsumed as data governance (Khatri & Brown, 2010), and technical capabilities, most importantly relating to enterprise-wide data integration and architecture (Ballou et al., 1998; Goodhue et al., 1988)

A third phase of data management in enterprises began in the 2010s with the use of *larger volumes of internal and external data (big data)* and the emergence of digital business models and data-driven services (Buhl et al., 2013; Provost & Fawcett, 2013; Wixom & Ross, 2017). These developments emphasize the business value and impacts of data (Chen, Chiang, & Storey, 2012; Clarke, 2016). The strategic role of data is reflected in additions to the data management-related knowledge base: The technological and organizational capabilities to acquire, store, and process the increasing variety and volume of data, based on data lakes and advanced analytics platforms (Abbasi, Sarker, & Chiang, 2016; Chen, Li, & Wang, 2015; O'Leary, 2014). Data management is also increasingly associated with strategic capabilities to enable data monetization by improving business processes and decision-making or by innovating business models (Chen et al., 2012; Schüritz et al., 2017; Wixom & Ross, 2017).

In sum, the role of data has evolved from an enabling resource to a strategic one. In response, data management has developed from a technological capability focused on single databases to an enterprise-wide organizational and strategic capability. This development is mirrored in the accumulation of data management-related knowledge, which required substantial adaptation and extension to cope with the evolving roles of data in businesses over time.

2.2 Knowledge Accumulation Challenges in Data Management

Despite the maturing body of data management knowledge, academics (Haug & Stentoft Arlbjørn, 2011; Marsh, 2005) and practitioners consistently report on the difficulties facing companies in managing data. Based on a review of empirical studies, (Marsh, 2005) summarized that "88 per cent of all data integration projects either fail completely or significantly over-run their budgets, … 33 per cent of organisations 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). These practical difficulties result from the sociotechnical nature of data management, and can only be solved by building strategic, organizational, and systems-related capabilities. Thus, data management can be framed as a "wicked" management problem (Rittel & Webber, 1973), i.e., a problem that addresses complex situations and is novel and unique, hard to define, and has no true-or-false solution.

Table 1. The Evolution of Data Management in Enterprises

	Phase 1: **Data administration** **(since the 1980s)**	**Phase 2:** **Quality-oriented data management** **(since the 1990s)**	**Phase 3:** **Extensions to strategic data management** **(since the 2010s)**
Business context			
Roles of data	• Data as a prerequisite for application development and as an enabler of automation in business functions	• Data as an enabler of enterprise-wide business processes and decision-making	• Data as an enabler of a firm's business models and value propositions
Data resources	• Databases for automated data processing in specific enterprise functions—for instance, accounting systems and inventory systems • Structured data	• Integrated information systems: enterprise resource planning systems (ERP), computer integrated manufacturing (CIM) • Data warehouses, business intelligence (BI) • Mainly internal, structured data	• Integrated and connected information systems • Data lakes and advanced analytics platforms • Large volumes of internal and external data (*big data*), comprising structured and nonstructured data sources
Data-related concerns	• Data model quality, data availability, data reuse (Gillenson, 1985)	• Enterprise-wide data integration, data quality (Goodhue et al., 1992; Grover & Teng, 1991; Ravindra, 1986)	• Business value and impacts, data compliance, data privacy, data security (Akter et al. 2016; Constantiou & Kallinikos, 2015; Xie et al., 2016)
Responsibilities for data	• Database administrator (Goldstein & McCririck, 1981; Weldon, 1981)	• Business process owners, later master data management (MDM) and business intelligence (BI) teams	• Chief data officer, data scientists, data analysts
Data management knowledge			
Perspective on data management	• Data administration (focus on databases)	• Quality-oriented data management (focus on data as an enterprise resource)	• Strategic data management (focus on data-driven innovation)
Management approach	• Database management	• Resource management, quality management	• Strategic management
Data management knowledge base	• Mainly database-related knowledge (data modeling) (Aiken et al., 2013)	• Data management-related body of knowledge, comprising organizational capabilities (i.e., data governance) (Khatri & Brown, 2010) and technical capabilities (i.e., data integration and architecture) (Ballou et al., 1998; Goodhue et al., 1988)	• Data management-related body of knowledge, extending the organizational and technical capabilities to acquire, store, and process the increasing variety and volume of data (Abbasi, Sarker, & Chiang, 2016; Chen, Li, & Wang, 2015; O'Leary, 2014). • Strategic capabilities to enable data monetization and data-driven innovation (Chen et al., 2012; Schüritz et al., 2017; Wixom & Ross, 2017)

Concerning knowledge accumulation, several challenges prevail: As an interdisciplinary field, data management draws on concepts and theories from various disciplines—most importantly, computer science (specifically databases and data analytics), information systems, and management. The knowledge base informing data management is created in both the research and the practitioner communities and the interactions between the two have led to the development of the most successful approaches—for instance, the *total data quality management* approach that transfers product quality management approaches to data management (Wang, 1998). To summarize, we argue that tackling wicked problems in data management requires the combination of knowledge across disciplines and from the research and practitioner communities.

2.3 Reference Models as Sources of Data Management Design Knowledge

Many companies are turning to reference models that should help them to build the strategic, organizational, and systems-related capabilities required for data management. In fact, substantial knowledge has been accumulated in the form of data management frameworks and reference models (Batini et al., 2009; Madnick et al., 2009). Based on a systematic review of practitioner and academic sources, we identified more than 10 data management reference models (see Table 2 for an overview and Table A1 in the Appendix for a detailed description) and many of these have a substantial and active base of contributors and users. The development of reference models is often the result of experts working together in industry-specific consortia or data management associations and synthesizing their practical experiences. Examples are the DAMA-DMBOK framework (DAMA, 2017), developed by the world's largest association of data management professionals, the EDM Council's data capability assessment model (EDM Council, 2018), developed by more than 200 companies and software vendors from the financial industries, the Performance Improvement Council's data quality maturity model (PIC, 2016), developed by 16 governmental agencies, the data quality management system (GS1, 2010), developed by the retail and consumer goods industry standardization body GS1, and the data governance maturity model (IBM Data Governance Council, 2007), developed by a software user group. Two reference models are from academic research, while the CDQM reference model (Hüner, Ofner, & Otto, 2009; Otto, 2011b) and its successor, the DXM, are the only models created through industry-research collaboration. The popularity of these frameworks that structure data management practices underpins not only the practical relevance and challenges of data management, but also reveal a substantial body of design knowledge that has been established via efforts by researchers and practitioners.

From an academic perspective, a reference model is a specific type of conceptual model (Frank et al., 2014; vom Brocke, 2007) that builds an abstract representation of domain knowledge relating to a selected phenomenon of interest. It facilitates understanding and communication among different stakeholders while supporting solution design, implementation, and maintenance (Wand & Weber, 2002). Reference models aggregate theoretical and empirical concepts, and have two key characteristics (Frank et al., 2014; Thomas, 2006; vom Brocke, 2007): their level of abstraction (i.e., they specify generally valid elements related to a phenomenon of interest) and their character as recommended practice (i.e., they serve as an orientation for designing company-specific models). Reference models result from design-oriented research (Feldman & Orlikowski, 2011; Goldkuhl, 2012), following the DSR methodological paradigm (vom Brocke & Buddendick, 2006). They can be created with both a descriptive and a prescriptive intention (Frank, 2007): On the one hand, they seek to provide substantial descriptions of a domain; on the other hand, they aim at "delivering blueprints for a distinctively good design of information systems and related organizational setting" (Frank, 2007, p. 119). Thus, they can be classified as descriptive and/or prescriptive knowledge, as described by Gregor's (2006) theory types 1 and 5. Prior research into reference models has taken either a use-oriented perspective that emphasizes their reuse and adaptation to create company-specific models (Thomas, 2007; vom Brocke, 2007), or a configuration-oriented perspective that focuses on configurative approaches in building reference models and their conceptual support in the form of configurable reference modeling languages (Becker, Delfmann, & Knackstedt, 2007; Recker et al., 2007).

Based on our analysis of the existing reference models in data management, we make three crucial observations:

(1) Reference models as a synthesis of descriptive and prescriptive knowledge: Reference models for data management synthesize knowledge in the form of conceptual, capability, and maturity models (see Table 2). Conceptual models are mostly the result of academic research and are expressed in the form of metamodels (in the case of CDQM/DXM) or classification models (in the case of the big data analytics capability model) (Gupta & George, 2016). Their role is to define and decompose data management, shedding light on the *what*. Capability and maturity models are often developed by practitioner communities to structure and assess data management practices, emphasizing the *how*.

(2) Independent development of academic and practitioner knowledge bases: With the exception of CDQM/DXM, we observe only very little interaction between practitioners and academics in the development of data management reference models. Our review reveals that existing reference models are dominated by practitioner contributions. While these are effective in synthesizing practitioner knowledge from local practices, they also have limitations. Their development process is not transparent, and they often lack consistency (owing to contributions from multiple authors) and clear ontological foundations. On the other hand, reference models that were created exclusively by academics risk having limited practical relevance. As they seek to contribute primarily to the scientific discourse, they tend to ignore the increasing amount of implicit knowledge embodied in concrete artifacts and local practices.

(3) Little knowledge accumulation over time: Only three reference models (i.e., CDQM/DXM, the DAMA-DMBOK framework, and the enterprise information management maturity model) accumulate data management knowledge over time, while the others have all been published once but never updated. With the exception of two (academic) frameworks targeted solely at big data management, most data management frameworks are still rooted in quality-oriented data management. This implies that they have not been revised to cope with the growing role of data (i.e., the move towards strategic data management) and thus risk becoming obsolete.

In sum, we find that in data management, reference models are important sources of design knowledge with a substantial and active base of contributors and users. However, the existing reference models and frameworks are ineffective at accumulating knowledge over time and integrating knowledge spread in academic and practitioner communities. This calls for a better understanding of how relevant, interdisciplinary knowledge is accumulated with reference models and highlights the need for guidelines to iteratively develop them.

3 Method

3.1 The Context and the Case Setting

To address our research question, we opted to study the case of CDQM/DXM and theorize on knowledge accumulation with reference models based on a retrospective analysis. This case study satisfies several criteria that justify a single-case study over a multiple case-design (Yin, 2014): (1) it represents a *longitudinal* DSR setting with strong research-industry collaboration over more than 12 years; (2) it is *revelatory*, since it provides a unique opportunity to observe and analyze the unexplored phenomenon of knowledge accumulation in reference models; (3) it is also an *extreme* case (Gerring, 2006) that is "prototypical or paradigmatic" of the phenomena of interest. It deviates from other reference models in that it accumulates knowledge in different versions of a reference model and in that DSR guidelines were used in their development.

The context of our study is an ongoing research program initiated in 2006 and has involved more than 30 companies and researchers from three universities in a *longitudinal DSR process*. These companies are large, Europe-based multinational enterprises, with annual revenues of more than €1 billion from different industries (including automotive, transportation, pharma, and consumer goods), thus supporting the generated knowledge's generalizability. They are typically represented by corporate middle-management roles with oversight over enterprise-wide data management practices, such as head of data management or enterprise architect. Thus, the company representatives contribute their experience with and vision of their firms' concrete data management approaches; most participated in the program for at least five years, ensuring continuity and allowing for knowledge accumulation and transfer.

As a form of *practice research* (Feldman & Orlikowski, 2011; Goldkuhl, 2012), the research program acknowledges the large body of knowledge in the scientific and the practitioner domains and *accumulates knowledge in close research-industry collaboration*. It relies on interplays between the subpractices of situational inquiry and theorizing in close research-industry interactions, similar to the ideas of collaborative practice research (Mathiassen, 2002). Its ultimate goal is to improve practices and to inform the research community, general practice, and local operational practice.

The research activities follow guidelines for design-oriented IS research (Hevner et al., 2004) and a rigorous iterative artifact design process in which researchers and practitioners define research objectives, assess the progress of work, and evaluate artifacts. The activities are systematically consolidated and have resulted in *different versions of a reference model*, starting with the initial version of the CDQM reference model (Hüner et al., 2009), its extensions, and its redesign in the form of the DXM (Pentek et al., 2017). Through our involvement in the research activities and their comprehensive documentation in working reports and academic publications, we have complete traces of the different stages and versions of artifact design.

Table 2. Reference Models for Data Management Knowledge Accumulation (Solution Space)

Phase*	Reference model	Reference model type	Origin
2, 3	DAMA-DMBOK framework – 1st edition (DAMA, 2009)	Capability model	Industry consortium (comprising data management experts in 53 active chapters on all continents)
	DAMA-DMBOK framework – 2nd edition (DAMA, 2017)	Capability model	
2, 3	CDQM reference model for corporate data quality management (Hüner et al., 2009)	Maturity and capability model	Research and industry consortium (30+ Europe-based companies, 3 universities)
	DXM data excellence model (Pentek, Legner, & Otto, 2017)	Maturity and capability model	
2	Master data management maturity model (Spruit & Pietzka, 2015)	Maturity model	Research (1 company for evaluation involved)
2	Data quality maturity model (PIC, 2016)	Maturity model	Industry consortium (US-based governmental agencies)
2	Data capability assessment model (EDM Council, 2018)	Maturity and capability model	Industry consortium (200+ companies from the financial sector)
2	Data governance maturity model (IBM Data Governance Council, 2007)	Maturity model	Industry consortium (52 companies, associations, and governmental agencies, 3 universities)
2	Data-quality management system (GS1, 2010)	Capability model	Standardization body (14 companies and 11 GS1 offices)
2	Master data quality management framework (International Organization for Standardization [ISO], 2011)	Capability model	Standardization body (no publicly available information on contributors to the reference model)
2	Data management capability model (Hopkins et al., 2018)	Capability model	Market analyst (no publicly available information on contributors to the reference model)
2	Enterprise information management maturity model (Gartner, 2014)	Maturity model	Market analyst (no publicly available information on contributors to the reference model)
3	Big data analytics capability model (Akter et al., 2016)	Capability model	Research (Delphi study with 51 and 43 respondents, survey with 152 responses)
3	Big data resources framework (Gupta & George, 2016)	Conceptual model	Research (2 studies with 232 and 108 responses from big data analysts and CDOs)

Notes: * Quality-oriented data management (Stage 2).
Strategic data management (Stage 3).

Table 3. The Consortium Research Activities and the Applied Techniques

	Analysis (problem space)	**Design (generic solution)**	**Evaluation (instantiations)**
Research activities			
General DSR activities	• Problem identification and motivation • definition of requirements and solution objectives	• Design and development	• Demonstration; evaluation
Corresponding activities in consortium research	• Exploring the problem space • creating a shared understanding of the phenomenon of interest (boundaries, rationale) • defining research objectives and requirements	• Reviewing academic knowledge base and emerging solutions to explicate (implicit) design knowledge • theorizing about design decisions and alternative solution designs; developing the metamodel and constructing the artifact	• Evaluating generic artifacts through expert feedback and focus groups (artificial evaluation) • instantiating artifacts in company settings (situational design) for demonstration and evaluation (naturalistic evaluation)
Techniques			
Plenary discussion: Presentation of research progress and results to an audience of >30 data management experts with the objective to build consensus.	• Review and confirmation of problem analysis and requirements	• Review and confirmation of design decisions and different versions of the artifact	-
Focus group: Working sessions (between 5 and 15 data management experts) moderated by a researcher to explicate implicit design knowledge and gather in-depth expert feedback.	• Exploration or confirmation of problems	• Review of emerging (situational and generic) solutions • discussion and confirmation of design decisions and artifacts	• Evaluation of (situational and generic) artifacts
Expert interviews: One-on-one interviews with subject matter experts from both the research and practitioner communities.	• Situational inquiry and problem identification	• Analysis of emerging (situational and generic) solution designs and artifacts	• Evaluation of (situational and generic) artifacts
Project: Projects involving researchers to instantiate and evaluate the artifacts to specific business settings.	-	-	• Instantiation of artifact (situational design) and evaluation
Case study: Qualitative research applied for the exploration and explanation of company-specific problems and solution designs.	• Situational inquiry and problem identification	• Examination of situational solution designs and artifacts	• Evaluation of (situational) artifacts
Survey: Data collection based on a semistructured or structured questionnaire.	• Confirmation of problems and requirements	-	• Evaluation of (generic and situational) artifacts
Desk research: Grounding the artifact design in the relevant scientific and practitioner knowledge base.	• Analysis of the scientific body of knowledge and the state of the art in industry	• Theorizing about design decisions and alternative solution designs • construction of the artifact	-

Specifically, the program applies the *consortium research* approach, as an organizational model for engaged research practice that follows design science guidelines (Back, Krogh, & Enkel, 2007). As a multilateral and longitudinal form of DSR, consortium research "aims at supporting and promoting collaboration between practitioners and academic researchers in a common area of interest in order to intensify the transfer of knowledge between these two groups" (Österle & Otto, 2010, p. 284). Consortium research typically unfolds in four activity categories (see Table 3) that reflect the DSR methodology (Peffers et al., 2007)—analysis (exploration of the problem space, leading to problem identification and the definition of requirements), design (development of the solution space via the iterative design and development of artifacts), demonstration and evaluation (via expert evaluation and situational instantiations), and diffusion (presentation and publication of the research results, targeted at general and local practice as well as the scientific community). These phases are conducted using a systematic and rigorous research approach by applying a specific set of techniques (Österle & Otto, 2010). Close interactions with practitioners are required to investigate situational designs (instantiations) and explicate (implicit) design knowledge and to review and confirm design decisions, ensuring the relevance, applicability, and utility of research results for the practitioner community. These interactions unfold in the form of plenary discussions (Österle & Otto, 2010) and focus groups (Tremblay, Hevner, & Berndt, 2010), expert interviews with subject matter experts (Meuser & Nagel, 2009), projects that instantiate the artifacts (Sein et al., 2011), case studies (Eisenhardt & Graebner, 2007), and surveys (Pinsonneault & Kraemer, 1993).

3.2 Data Collection and Analysis

To analyze the research activities and results, we followed Miles, Huberman, and Saldaña (2014) for qualitative data analysis, applying event listing, conceptual clustering, case analysis meeting, and case dynamics techniques. As the authors were involved in the research activities, we had firsthand knowledge about the work conducted in the consortium and the resulting artifacts over more than 12 years. We started by collecting and reviewing relevant sources that documented the knowledge accumulation throughout the research program. These sources included agendas and presentations from 61 consortium workshops, field notes from 41 focus groups and plenary discussions, transcripts and field notes from expert interviews, and research project documentations. Further, we reviewed 12 doctoral theses, nine case study reports, and multiple conference and journal publications that documented the research program's activities.

We applied inductive reasoning (Gregor, Müller, & Seidel, 2013) to trace how knowledge has been accumulated throughout the longitudinal DSR process from two complementary perspectives: the research process and the artifact design and evolution.

Research process analysis: We compiled an *event-listing matrix* (Miles et al., 2014) chronologically documenting all activities of the 12-year research program. Two researchers independently reviewed and coded the data. In line with the DSR literature (Gregor & Hevner, 2013), each research activity was classified according to its contributions to the problem space (i.e., nature, boundaries, rationale, and implementation) and the solution space (i.e., reference model design and reference model instantiation), resulting in a *conceptually clustered matrix*. In a *case analysis meeting*, the two researchers discussed deviating evaluations and jointly decided on a common classification. Finally, they identified and discussed highly relevant activities in the research process, in which new aspects of the problem space or additions to the solution space emerged, considerations arose, or important decisions were made, and marked these activities as decisive events.

Artifact design and evolution analysis: For the different versions of the reference model (i.e., the CDQM and the DXM), we systematically analyzed and compared the structure and contents of the artifact based on the relevant documents and publications in a *case dynamics matrix*. For this purpose, we used a coding scheme that reflects DSR concepts, namely metarequirements, design decisions, and design areas (the metamodel). To trace how the artifact design evolved over time, we identified the changes encountered during the course of the research process. Starting with the definition of a design area (typically, this implies setting the boundaries and defining the key objects), we observed that changes materialized in the form of refinements (as a result of the analysis and the comparison of alternative practices, their results, and the definition of key principles), extensions that broadened the scope of the design area, and modifications to the design area (typically seeking to either improve the design or correct inconsistencies or errors without refining it).

4 The Accumulation of Design Knowledge Throughout the Longitudinal Research Process

Over the course of the 12-year research program, close practitioner interactions in 61 workshops and research activities in 12 doctoral dissertation projects contributed to the evolution and accumulation of knowledge in different versions of a reference model for data management. Knowledge accumulation can be traced throughout our longitudinal DSR process, by analyzing decisive events (DEs) and activities that shaped the nature, boundaries, and rationale that underlie artifact design (i.e., the problem space) and the different versions of the artifact (i.e., the solution space). During the research process, we were able to identify 45 events that were decisive, since they either culminated in an understanding of the problem space or informed the design of the solution space (for details, see Figure 1 and Table A2 in the Appendix).

Based on this analysis, we found that knowledge accumulation occurred in stages as a result of maturing abstract and situational domain knowledge (solution space), and in response to the evolving roles of data (problem space). From a process perspective, knowledge accumulation materializes in three phases (see Table 4): (1) framing the problem and creating a shared understanding about enterprise-wide data management (*ontology*), (2) assessing maturity and building the required data management capabilities (*capability building*), and (3) addressing the growing data requirements of a digital and data-driven enterprise (*reorientation*). We explain the three phases below.

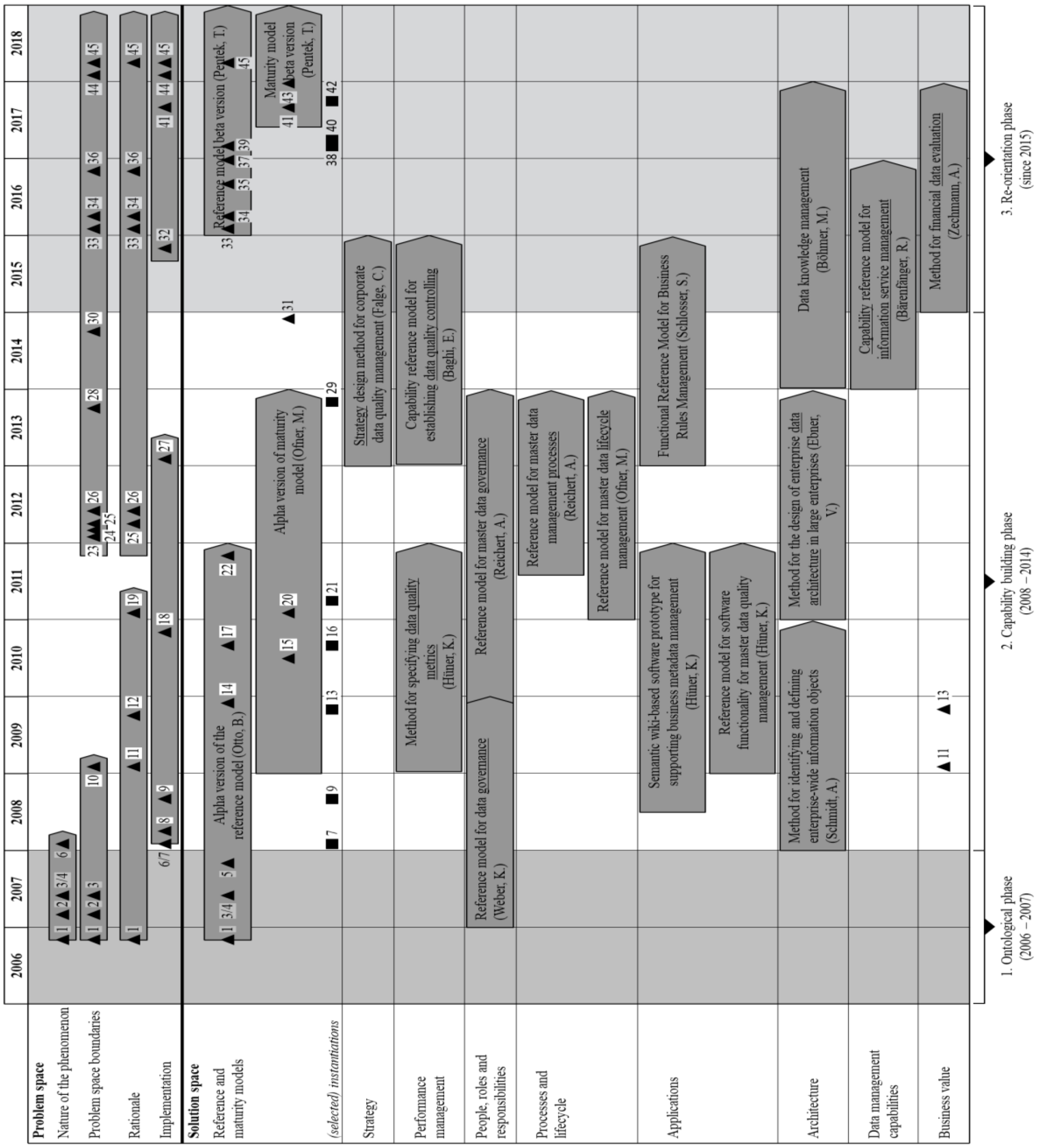

Figure 1. The Research Process and Decisive Events (DE) During the Longitudinal DSR Process

Table 4. The Knowledge Accumulation Phases

	Phase 1: Ontology	Phase 2: Capability building	Phase 3: Reorientation
Time	2006 to 2007	2008 to 2014	Since 2015
The problem space			
Research questions	• What is enterprise-wide data management? • What are its constituents (nature, boundaries, rationale)?	• How does one build enterprise-wide data management capabilities? • How does one assess the data management's maturity (implementation)?	• What is data management for digital and data-driven enterprises? • What are changes to the artifacts from Phases 1 and 2 (boundaries, rationale, implementation)?
Boundaries	• Master data	• Master data	• All data types
Nature	• Enterprise-wide (quality-oriented) data management	• Enterprise-wide (quality-oriented) data management	• Enterprise-wide (strategic) data management
Rationale	• Data quality	• Data quality	• Data excellence
Implementation	-	• Data management capabilities and maturity levels	• Continuous improvement
The solution space			
Artifacts	• Initial version of the reference model (alpha version of generic artifact)	• Refinement of the reference model: maturity assessment and refinement of design areas via methods, tools, and guidelines	• Reorientation and revision of the reference model (beta version of generic artifact): modification and extension of maturity assessment and design areas via methods, tools, and guidelines
Instantiations	• Mainly expert feedback (artificial evaluation methods), or explication of emerging situational design	• Company-specific instantiations (situational design)	• Company-specific instantiations (situational design), or explication of emerging situational design

4.1 Phase 1: An Ontology for Quality-Oriented Data Management

The research activities began in 2006 with the formation of a research consortium by a small group of data management experts from practice and academia. Although the user companies had been administrating huge amounts of data for decades, they experienced significant data quality issues owing to a lacking enterprise-wide perspective on data and a growing number and complexity of silo applications. Following the consortium research approach, the research activities began with an analysis of practical problems resulting from poor data quality, but—more generally—centered around fundamental questions relating to the phenomenon of interest (DE 1 and 2):

> **Nature:** What are the constituents of enterprise-wide (quality-oriented) data management?
>
> **Boundaries:** What should be considered part of enterprise-wide (quality-oriented) data management?
>
> **Rationale:** What are the issues resulting from poor data quality? How can data quality be improved?

To answer these questions, the research activities sought to develop a reference model for quality-oriented data management. Companies had been asking for such a model, which they could use as an orientation for defining their company-specific approach, since they were unsure about how to approach and institutionalize data management in their organizations. The reference model would also help them to educate employees and to communicate the required approach to the large number of stakeholders that contribute to data management initiatives.

Over two years, from 2006 to 2007, four focus groups and five plenary discussions were conducted to explore the problem space, define the objectives, and discuss design decisions and the initial version of the reference model. The reference model design started with a comparison of these questions to state-of-the-art concepts in academia and practice (DE 4). Based on the practitioners' experiences and by using business engineering (Österle & Winter, 2003) as the conceptual foundation, the reference model for CDQM emerged as the alpha version of the artifact (DE 5). This model is strongly rooted in the tradition of quality-oriented data management and focuses on quality-assured master data that mainly contributes to error-free business process execution (e.g., deliveries or invoicing) and to meaningful reporting (e.g., enterprise-wide sales reporting).

4.2 Phase 2: Capability Building for Quality-Oriented Data Management

Over time, an increasing number of companies started using and instantiating the alpha version of the reference model to design, implement, and communicate their data management approach. Given companies' increasing experience with the artifact and with data management generally, the research activities centered around more detailed aspects and implementation questions (DE 6 and 8):

> **Implementation:** How does one build (quality-oriented) data management capabilities? How does one determine and assess different maturity levels?

Building on the reference model for CDQM from Phase 1 and feedback from the companies, the research activities in Phase 2 aimed at detailing the model to provide methodological guidance for building data management capabilities. From 2008 to 2014, the six design areas of the CDQM reference model were further refined by methods, architectures, and tools developed in nine dissertation projects. As one of the first research activities, data management boards and roles were defined (DE 12), resulting in a reference model for data governance (Weber, Otto, & Österle, 2009; Wende, 2007). These roles and their responsibilities were later further detailed by Otto and Reichert (2010) and were complemented by master data management processes (Reichert, Otto, & Österle, 2013). The researchers developed granular data quality metrics (Hüner et al., 2011; Otto, Ebner, & Hüner, 2010; Otto, Hüner, & Österle, 2009) as well as an overarching capability reference model for data-quality control (Baghi, Otto, & Österle, 2013). Further research activities resulted in reference models for data application functionalities (Otto, Hüner, & Österle, 2012) and methods for data architecture (Baghi et al., 2014). To address the need to monitor and benchmark the progress of data management, the consortium decided to complement the reference model with a more detailed view on the required practices and maturity levels (DE 6 and 7) for each design area. The design activities, which covered a period of five years, included several iterations of design and evaluation as well as intensive collaboration with practitioners (DE 15, 16, 20, and 21). They resulted in a maturity model that was also adopted by the European Foundation for Quality Management (EFQM) as a recommended approach for quality-oriented data management (EFQM, 2011).

4.3 Phase 3: Reorientation toward Strategic Data Management

From 2012 on, the consortium discussed the increasing strategic relevance of the data resource (see Table 2), which not only improves internal business processes and decision-making, but also enriches the external value propositions (DE 23 to 28 and 30). However, it was only in 2015 that the consortium realized that these developments would fundamentally impact data management. In that year, we decided that the CDQM reference model should be revised with the goal of supporting companies in their transformation toward digital and data-driven enterprises (DE 32). Research activities in Phase 3 started by analyzing the requirements (DE 33, 34, and 36). This included reconsidering fundamental questions about enterprise-wide data management, but in a broader context—mirroring the considerations in Phases 1 and 2:

> **Boundaries:** What should be considered as part of strategic data management in digital and data-driven enterprises?
>
> **Rationale:** How can data be used to create and maximize business value?
>
> **Implementation:** How does one build data management capabilities in digital and data-driven enterprises? How does one assess the maturity of data management in digital and data-driven enterprises?

This phase is ongoing; it comprises four dissertation projects and has resulted in a beta version of the reference model for data management (DE 33 to 35, 37, and 39). This also led to the maturity model being updated (DE 41 to 43).

5 The Accumulation of Design Knowledge with Different Versions of the Reference Model

Knowledge accumulation materialized not only in decisive events during the 12-year research process, but also in the reference model's different versions and changes to its structure and content (see Table 5).

Table 5. The Accumulation of Design Knowledge in the Reference Model's Development

	Phase 1: Ontology	Phase 2: Capability building	Phase 3: Reorientation
The problem space: Metarequirements			
Metarequirements	R1.1: Outline key constituents of data management (nature) R1.2: Consider enterprise-wide master data as the most relevant data type (boundaries) R1.3: Improve data quality (as rationale for data management)	R2.1: Develop data management in stages (implementation) R2.2 Assess the maturity of data management (implementation)	R3.1: Identify business-critical data needs and address relevant data-related concerns (rationale—extended) R3.2: Manage data from different sources and for different purposes (boundaries—extended) R3.3: Demonstrate the value contribution of data management to the business (nature—extended)
The solution space: Artifact design			
Design decisions	DD1: Explicate a data strategy DD2: Develop data management capabilities through governance and system-related aspects	DD3: Understand data management as a continuous improvement cycle	DD4: Translate business capabilities into data management capabilities DD5: Manage the data lifecycle DD6: Demonstrate results in terms of data excellence and business value
Artifact characteristics	The CDQM as conceptual model: Definition of six design areas	The CDQM as capability and maturity model: Refinement of 5 and modification of 1 design area: principles and practices that support the design goals	The DXM as capability model and maturity model: 6 extended, 5 new design areas
Goals: Business capabilities	-	-	D
Goals: Data management capabilities	-	-	D, R
Goals: Data strategy	D	R	E
Enablers: People, roles, and responsibilities	D, R	M	E
Enablers: Processes and methods	D	R	E
Enablers: Performance management	D	R	M, E
Enablers: Data lifecycle	-	-	D, R
Enablers: Data architecture	D	R	E
Enablers: Data applications	D	R	E
Results: Data excellence	-	-	D, R
Results: Business value	-	-	D, R

Notes:
D = definition (setting the boundaries and defining the key objects)
R = refinement (analyzing and defining practices, results, and principles)
E = extension (broadening the scope)
M = modification (improving/changing/correcting)

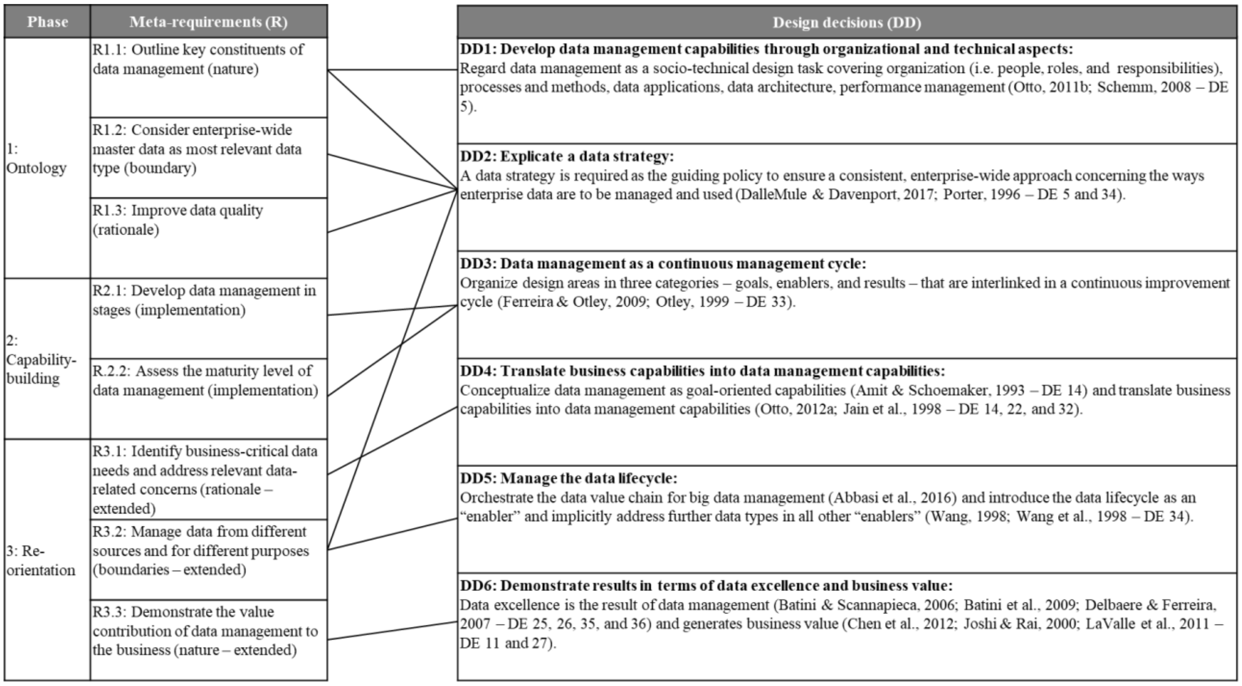

Figure 2. The Metarequirements and the Design Decisions

5.1 The Reference Model's Purpose and Scope

Building on the understanding of data as an economic good and the resource-based view (RBV), the CDQM and the DXM, as different versions of the reference model, had a shared purpose: to help organizations manage data as a strategic resource. Concerning scope, both address global corporations, which typically have complex organizational structures and distributed operations, resulting in data silos and a lack of transparency concerning the data resources. In these corporations, the challenges of establishing enterprise-wide data management are particularly salient, since the complexity of data-related issues and data processing increases when an organization and its systems are more distributed (Jain et al., 1998). Implementing enterprise-wide data management requires significant changes to existing policies and practices and impacts on headquarters, business lines, and every subsidiary (Haug & Stentoft Arlbjørn, 2011).

5.2 The Artifact Evolution During the Three Phases

The different versions of the artifact reflect design knowledge accumulation through changes to the problem space (i.e., the metarequirements) and the solution design (i.e., the design decisions and the artifact's structure and content) (see Table 5). The current version of the reference model—the DXM—reflects six key design decisions (DD) that address eight metarequirements (see Figure 2). These design decisions and metarequirements represent the evolving problem and solution spaces in the three phases. The structure and content of the artifact evolved around 11 design areas, from the *design area's definition* (setting the boundaries and defining the key objects), to *refinement* (analyzing and defining practices, results, and principles), to *extension* (broadening the scope) and *modification* (improving/changing/ correcting).

From Phase 1, *ontology building*, three key requirements were formulated, with the goal of clarifying the understanding of the phenomenon of interest. First, the reference model should outline the key constituents of enterprise-wide data management, i.e., specify strategic, organizational, and technical aspects that are relevant for managing data (R1.1 Nature). Second, the artifact should set the scope on master data as the company's key resource. Master data is considered most critical, since it refers to an organization's core business entities (Smith & McKeen, 2008; Dreibelbis et al., 2008) and is often referenced in business processes and reports. Thus, the focus of the reference model was on typical master data classes, such as product and material master data, supplier and customer master data, and master data regarding employees, assets, and organizational units (R1.2 Boundary). Third, and in line with the evolution of data management, improved data quality was considered the main goal and result of data management. Thus, the artifact should help to improve data quality (R.1.3 Rationale). Based on these requirements, the CDQM as alpha version sought to provide a reference model for enterprise-wide data management (the nature of the artifact) with the goal of improving data quality (the rationale) and focusing on master data (the boundaries). The development of the artifact was guided by two design decisions. First, the dual nature of data management as a sociotechnical design task resulted in the design decision to consider both organizational and technical aspects (DE 5; DD1). Second, the data strategy (DE 5 and 34; DD2) should clarify the role and define the guiding principles for the enterprise-wide data. The alpha version of the reference model (Otto, 2011b; Schemm, 2008) provides a shared terminology and decomposes quality-oriented data management into six design areas (strategy, controlling, organization, processes, architecture, and applications).

In Phase 2, *capability building*, companies realized, as a result of the artifact's instantiation in practice, that it took them several years to address and implement the design areas at an enterprise-wide level. They built their data management capabilities very slowly, owing to the numerous stakeholders involved and the complexity of organizational and technical changes. This resulted in two requirements: to develop data management capabilities in stages (R2.1) and to assess the data management maturity level (R2.2). The key design decision in this phase was to reflect a management cycle with a staged development and continuous improvement cycle (DE 33; DD3). Thus, the reference model evolved from providing the ontological foundations to addressing questions of implementation. This introduced the need to refine the artifact's design areas in order to explain capability building and to distinguish different practices and maturity levels. As a maturity model, the reference model details each design area and comprises—at its most detailed level, 30 practices and 56 measures. It can be used as a concrete assessment element during an appraisal (Ofner, Otto, & Österle, 2013).

In Phase 3 (the last and ongoing *reorientation phase*), the emergence of digital and data-driven business models placed new requirements on data management and shifted the focus from data quality to business value from data. R3.1 addresses the growing business criticality of data in digital and data-driven enterprises. Identifying and addressing a company's data needs requires, besides technical and organizational capabilities, close alignment between data management and the business's strategic goals. It also requires mitigating data-related risks and complying with an increasing number of regulations, relating, for instance, to data privacy or traceability. R3.2 refers to the growing number of data sources and the volume of available data, such as smart factories, smart products, or social media. To make use of big data and to generate data-driven insights, data management must expand its traditional scope concerning master data to include all relevant data types, including analytical, web, or sensor data.

Finally, in light of the digital and data-driven economy, the value generated by data and data management's contributions to the business activities must be transparent (R3.3). To address these requirements, a major revision of the reference model was necessary. The design decision was taken to incorporate an outside-in perspective in the DXM. The DXM should translate business capabilities into data management capabilities, emphasizing that data management is contingent on business objectives (DE 14, 22, and 32; DD4). In line with the management of physical resources, data management assures an enterprise-wide consistent approach to create, maintain, use, and archive data (DE 34; DD5). As an outcomes-oriented capability, data management contributes to two outcome types: data excellence and business value (DD6). First, data management has direct impacts on data qualities, defined in the reference model as data excellence (DE 35). Second, data excellence creates value to the business, reflected by the business value design area (DE 25, 26, and 36).

5.3 The Data Excellence Model as the Beta Artifact

The beta artifact and current version, the DXM, is a reference model for data management in which data management comprises the strategic, organizational, and technological capabilities necessary to deploy data resources in a way that creates business value. It builds on the understanding of data management as a set of capabilities that are contingent on business objectives and that materialize in the DXM structure:

- Capabilities are *goal oriented* (Amit & Schoemaker, 1993). Accordingly, the goal of the DXM goal is to support the business capabilities and to define the data management capabilities required for their support.
- Capabilities are provided by a *resources and abilities set* (Stoel & Muhanna, 2009). The enablers section specifies the sociotechnical design areas for providing the required data management capabilities.
- Capabilities are *outcomes oriented* (Bharadwaj et al., 2013). Data management capabilities seek to maximize business value through data excellence (as results).

Figure 3 depicts the DXM and its 11 design areas, which represent the main constituents (or domains) of data management and are further described in Table A3 (in the Appendix). Each of the design areas is ontologically defined through the entities (or constructs) it addresses and through result documents that represent the outcomes of the design activities.

The constructs and their relationships are specified in the form of a metamodel (i.e., a conceptual data model of the domain), to build the ontological foundation and to create a shared understanding among experts from academia and practice (for the comprehensive metamodel that covers all the design areas, with their constructs and relationships, see Figure A1 in the Appendix). To support capability building and continuous improvement, the DXM further refines each design area by a set of practices and principles that ensure that the design goals materialize. It is supported by justificatory knowledge from the scientific and practitioner domains.

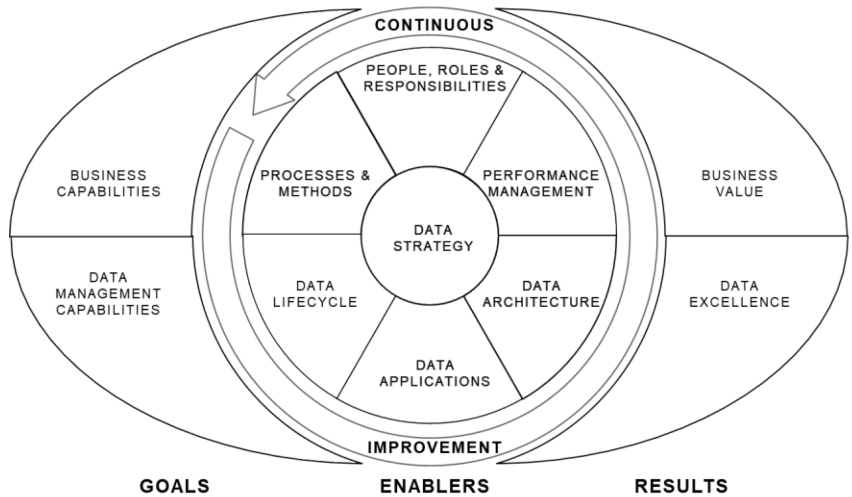

Figure 3. The Data Excellence Model (DXM)

5.4 The Application of the Data Excellence Model

The CDQM and DXM were adopted by hundreds of enterprises[1] inside and outside the research consortium, proving the design areas' validity as well as the reference model's applicability and usefulness. As part of the research activities, several rounds of demonstration and artificial and naturalistic evaluation (Venable, Pries-Heje, & Baskerville, 2016) were conducted. These allowed for identifying typical scenarios for applying the reference model and thereby reusing design knowledge (see Table 6). They can be categorized as (1) translating the abstract design knowledge into concrete situational designs (instantiation), or (2) using the reference model as abstract situational knowledge for assessment, communication, and education purposes (mobilization).

Instantiation: **from abstract design knowledge to company-specific situational design.** This scenario corresponds to the established understanding of reference models from the literature, i.e., their role as blueprint for company-specific instantiation. The results are situational designs (Goldkuhl, 2011; Winter, 2008), which, in our study, may be either prescriptive (the to-be situation in cases A, B, and C) or descriptive (explicating an existing or emerging design, cases D and E). Tailoring to the situational model may include company-specific refinement (e.g., defining company standards) and involves renaming design areas and adjusting layout (colors, symbols) to comply with corporate guidelines (cases A, B, and C). With the extended scope of data management, companies increasingly apply the reference model to analyze approaches to manage new data domains. Thus, the reference model serves as a diagnostic tool in digitalization or big data initiatives to explicate emerging data management capabilities (cases D and E).

Mobilization: **relating internal activities to abstract design knowledge.** The reference model is also often applied as a communication and educational tool that synthesizes generic design knowledge from the perspective of a firm. It supports communication, for instance, in the employee magazine and intranet, and motivates internal data management activities by referring to an established and legitimized body of design knowledge (case F). Also, both versions of the reference model were used to structure and develop education programs for data managers (case G). Finally, as shown in case H, companies use the generic artifact to assess and benchmark their data management initiative (with the reference model as the underlying domain model).

[1] Since the alpha and beta versions of the artifact are public and form the basis of recurrent training programs, it is hard to know the exact number of users. Our estimation is based on the cumulated number of companies participating in the consortium, additional consulting projects, and participants in CDQM/DXM training sessions.

Table 6. The Application of the Reference Model (Examples from Phase 3)

Scenarios		Illustration
1. Instantiation: Generic artifact (abstract design knowledge) to concrete situational design	1.1 The reference model applied and adapted for developing concrete situational models	• Case A: A pharmaceutical company developed a data strategy based on the beta version of the reference model. The strategy was detailed by defining activities for improving each design area of the enablers. • Cases B & C: A pharmaceutical company and a transportation company developed and issued a corporate policy for data management based on the beta version of the reference model. These policies rely on the reference model to introduce definitions, outline the design areas, and instantiate each enabler via relevant standards.
	1.2 The reference model applied and adapted for explicating and analyzing (emergent) concrete situational models	• Case D: An automotive supplier analyzed its current data management practices for conditions-based monitoring using the reference model. The goal was to develop a shared understanding of data management in this new domain among different stakeholders. • Case E: A focus group comprising 15 experts from 11 companies analyzed their data management challenges and practices for data lakes. The reference model allowed them to document emerging and required practices.
2. Mobilization: Generic artifact (abstract design knowledge) for internal mobilization	2.1 The reference model as a communication and education tool	• Case F: A consumer goods company leveraged the reference model to inform its employees about data management and to communicate internally. • Case G: The generic artifact has been used to build a training program for data management teams. To date, five programs have been conducted based on the alpha version of the artifact. The program is currently being revised to align with the beta version of the artifact.
	2.2 The reference model as a basis for maturity assessment/benchmarking	• Case H: A transportation company regularly assesses its data management activities' maturity to monitor progress and to identify improvements.

6 Reflections on Design Knowledge Accumulation Mechanisms

By analyzing a revelatory and extreme case of iterative reference model development over more than 12 years, we were able to theorize on the fundamental logic of longitudinal, multilateral design knowledge accumulation. As a synthesis, we have derived knowledge accumulation mechanisms from both the process perspective and the artifact perspectives.

6.1 The Process Perspective: Knowledge Accumulation Stages

Our case insights reveal that knowledge accumulation materializes in response to wicked problems that are triggered by technology's evolving roles in businesses (problem space) and results from maturing abstract and situational design knowledge (solution space). We suggest distinguishing three knowledge accumulation stages according to the dimensions of problem domain maturity and solution domain maturity (see Table 7):

Reference model as ontology: Whenever technology's evolving roles create complex sociotechnical challenges (*wicked problems*), the initial stage is characterized by low experience and little expertise with the problem domain as well as low maturity of the available knowledge and solutions. Since different stakeholders are involved and often have vague ideas, the reference model helps to reduce ambiguity in the problem interpretation and in the understanding of the problem and solution spaces. From a DSR perspective, this goes hand in hand with setting a suitable scope, defining the boundaries of the phenomenon of interest, and clarifying the rationale underlying the solution design. Through the observation and explanation of the emergent phenomenon, descriptive knowledge about the problem and solution spaces is created. Practitioners directly experience the problems and often experiment with first solution ideas. Academics analyze these practical experiences and review the knowledge base for theories that help to explain the emerging phenomena but also seek prior conceptualization that can be reused, refined, or extended for solution design. Challenges during this stage arise from different disciplinary vocabularies and concepts, a lack of shared semantics, and difficulties in structuring the problem space. In this stage, the reference model proved to help clarify ontological foundations as a conceptual model that acts as a shared language and a foundation for integrating different stakeholders' worldviews. Expressed as a metamodel, it defines shared semantics and helps to decompose the phenomenon of interest in design areas.

Table 7. The Stages of Design Knowledge Accumulation with Reference Models

	Phase 1: Ontology	**Phase 2: Capability building**	**Phase 3: Reorientation**
Sociotechnical phenomenon			
Research trigger	• Observation of a complex sociotechnical phenomenon (wicked problem), triggered by the evolving roles of technology in businesses	• Questions related to implementation, capabilities, and maturity	• Observation of fundamental changes relating to the phenomenon of interest • Questions related to implementation, capabilities, and maturity
Maturity of the problem domain	*Emerging (problems are fuzzy and not well understood)* • Lack of shared understanding • Low expertise and little experience	*Maturing (a good understanding of the problem domain)* • Increasing expertise and experience • Fragmented body of practical and academic design knowledge	*Emerging (problems as a result of technological progress and extended technology use)* • Lack of shared understanding • Low expertise and little experience
Maturity of the knowledge and of the solution domain	*Low (no or poorly developed knowledge base)* • Lacking an ontological foundation • Scattered approaches and knowledge sources from different disciplines, but poorly integrated	*Maturing (increasing body of knowledge)* • Ontological foundations exist, but lacking knowledge about capability building	*Low (need for revision and extension of the knowledge base)* • Ontological foundations need to be recreated • Adaptation and extension of the knowledge base
The reference model			
Artifacts	Conceptual model: • Initial version of the reference model (alpha version of generic artifact)	Capability and maturity model: • Refinement of the reference model: maturity assessment, refinement of design areas via methods, tools, and guidelines	Revised conceptual, capability, and maturity model: • Reorientation and revision of the reference model (beta version of the generic artifact): the modification and extension of maturity assessment and design areas via methods, tools, and guidelines
The reference model's contributions	• Consistent and shared representation of the domain • Aggregate and integrate knowledge across domain boundaries	• Elaborate organizational practices • Plan the establishment of capability • Integrate the solution space	• Connect the emerging problem space to a theoretical basis • Integrate the solution space
DSR activities supporting knowledge accumulation			
Key activities	• Framing the problem and setting scope (nature, boundaries, and rationale) • Creating a shared understanding and ontological foundations (metamodel) • Decomposing the phenomenon of interest	• Establish a knowledge base building on experience and expertise from academics and practitioners • Support capability building by refining the artifact and each design area (practices, principles, and maturity levels)	• Revising scope (nature, boundaries, and rationale) • Adapting and extending the ontological foundations (metamodel) • Integrating new knowledge assets
Challenges	• Integrating different disciplinary vocabulary and concepts • Reducing problem ambiguity and structuring the problem space (wicked problem)	• Integrating fragmented practices • Reflecting situativity	• Integrating two evolutionary stages • Compatibility across stages

Reference model for capability building: This stage is characterized by a maturing solution space for the given problem, building on the ontological foundations, and by a growing yet fragmented body of practical and academic design knowledge. In this stage, questions relating to implementation, capability building, and maturity are raised. As a response, the reference model helps to explicate, integrate, and consolidate the fragmented design knowledge that is available in the form of situational designs (typically in the form of company-specific artifacts) and emerging practices (advocated by experts from practice and research). To create relevant artifacts, the reference model should also explicate and integrate practitioner knowledge about situational designs, which is often tacit (i.e., possessed by individuals, and not systematically documented). In this stage, knowledge accumulation materializes in DSR activities that detail each design area and analyze and derive principles and practices from practitioner and academic knowledge based on the ontological structure defined in Phase 1. Thus, the reference model evolves into a prescriptive model about the solution space that integrates different practices and maturity levels with the goal of capability building.

Reorientation to cope with fundamental changes: When an emerging domain matures and becomes a well-established domain in both practice and academia, the underlying design knowledge should stabilize. However, technological progress may lead to changes in the problem space that interrogate the principles and fundamental assumptions of artifact design and lower the projectability of design knowledge. From a DSR perspective, reference models are key for knowledge reuse after these fundamental changes in the problem space, but this phase requires that one reconsider the scope, the boundaries of the phenomenon of interest, and the rationale underlying the artifact design. Thus, this stage must recreate the ontological foundations of and add lacking aspects to the artifact design while revising and extending the existing knowledge base. Our experiences reveal several challenges and difficulties with updating the existing design knowledge in the reorientation phase. These include the reuse, revision, and extension of the existing knowledge base, as well as possible tensions between contributors and users of existing reference models and other groups that promote digital innovations. While contributor try to protect and extend their knowledge base, users tend to emphasize a phenomenon of interest's novelty and may question the utility of extending established conceptualizations.

6.2 The Artifact Perspective: Reference Models as Design Boundary Objects

Our study emphasizes the role of reference models as design boundary objects. Reference models help to explicate, integrate, and accumulate design knowledge alongside an evolutionary change in the problem space and solution space between different communities of research and practice and between general and specific situational requirements and solutions. Thus, reference models fulfill several functions that Bergman, Lyytinen, and Mark (2007) associate with design boundary objects: they promote shared representations, transform design knowledge, mobilize for design action, and legitimate design knowledge. For multilateral, interdisciplinary DSR settings, we suggest three roles of a reference model as a design boundary object:

First, the reference model *fosters a shared understanding by decomposing the phenomenon of interest and integrating design knowledge from different disciplines.* It proposes a representational model that outlines the key constructs and their dependencies via the definition of design areas. Thus, it aligns the relevant views on the phenomenon of interest and ensures consistency between technical, organizational, and strategic design choices. It thereby integrates technical and behavioral knowledge relating to the sociotechnical phenomenon, as suggested by Niederman and March (2012). By adding a visual to the formal representation, reference models become more accessible for practitioners.

Second, the reference model acts as a *boundary object between general and specific situational requirements and solutions.* As generic and abstract design knowledge, the reference model explicates (implicit) design knowledge from situational inquiry and materialized instantiations, but also forms the basis for creating company-specific situational designs (instantiation) and for assessment, education, and communication purposes (mobilization).

Third, the reference model functions as a *boundary object over time,* and its design reflects *the evolutionary nature of both the problem space and the solution space.* Once a shared and conceptually consistent understanding is established, the capability building phase assesses concrete organizational practices, emerging concepts and principles, and their utility in meeting the design goals. Accordingly, the reference model evolves from an artifact that organizes descriptive (conceptual) knowledge about problem spaces and solution spaces into an artifact that structures prescriptive knowledge in the form of a capability or maturity model.

Finally, to be effective, the reference model must cope with fundamental changes in the problem space and the solution space triggered by the development of new technologies. A constant cycle of DSR activities that underlie an artifact's design, as followed, for instance, in the consortium research method, help to reorientate the reference model toward continuous practical utility and to meet the new requirements. Comparing this development to competing approaches, one can observe significant differences in the way that new knowledge is integrated in phases of reorientation. The DAMA DMBOK framework, for instance, reflects the changing roles of data in the transition from its first version in 2009 to its second version in 2017; however, new constructs were simply added to the model without revising its conceptual structure. Based on our experiences, the reorientation of a reference model requires one to redefine the scope and integrate new theoretical and practical knowledge. This clearly goes beyond merely adding new objectives and design areas to an existing model. If such change is not addressed, the knowledge accumulated in the reference model risks becoming inadequate.

7 Discussion, Implications, and Limitations

7.1 Discussion

We have analyzed an extreme case of an interdisciplinary, multilateral DSR setting that reveals the fundamental logic of knowledge accumulation with reference models in a close research-industry collaboration. Our analysis of reference models in data management reveals a significant gap between the academic discourse, which has focused on reference model reuse and configuration (Frank et al., 2014) and the pragmatic reference model development processes driven by practitioner communities. Against this backdrop, our study has demonstrated how reference modeling, DSR, and practice research (Goldkuhl, 2011) complement one another to create relevant artifacts that accumulate design knowledge for wicked management problems. Based on our experience and our comparison to competing artifacts (see Table A4 in the Appendix), we have found that multilateral DSR settings with research-practitioner collaborations contribute to accumulating and integrating interdisciplinary design knowledge and are more likely to produce comprehensive and consistent reference models. We see three main contributions from researchers participating in reference model development: From a methodological perspective, following DSR guidelines for the systematic development of artifacts in multilateral settings results in more consistent results than practitioner contributions. From a content perspective, researchers contribute relevant academic concepts and theories that help frame the problem and solution spaces and develop theoretically grounded artifacts. While practitioners are good at providing their ideas and experiences with specific practices and their knowledge of situational designs, most of their design knowledge is tacit. Researchers can support practitioners throughout the process in explicating the concepts and design principles inherent in situational designs and integrating this knowledge into the reference model. Beyond their contributions to problem framing and artifact development, researchers also play a key role in evaluating the utility of generic and situational artifacts and in reflecting on the situativity and the conditions under which they work.

From our study on data management, we conclude that reference models, as sources of descriptive and prescriptive knowledge, have key roles in tackling wicked problems associated with increasing digitalization. Our longitudinal research process simultaneously reveals that the so-called “third wave of IT-driven competition” (Porter & Heppelmann, 2015), with its emphasis on big data and advanced analytics, does not cause groundbreaking shifts in data management, but should be considered as an opportunity to revise and extend traditional capabilities. For data management, we find that fundamental design principles such as the explication of a data strategy or the dual aspects of organizational and system-related capabilities remain relevant. Our findings also reveal that existing design knowledge needs to be constantly updated, extended, and revised in light of digital technologies’ increasing business criticality.

Finally, our study also offers a methodological contribution, since it outlines an approach to analyze knowledge accumulation in different stages and versions of DSR artifacts. We suggest tracing knowledge accumulation from a process perspective by analyzing decisive events that frame the problem space (i.e., decisions about the nature, boundaries, and rationale of the artifact) and shape the solution space (i.e., decisions related to the design). We complement this perspective with an artifact perspective that traces the evolving artifact structure and content based on DSR concepts, from definition to refinement to extension and/or modification. Given the lack of methodological guidelines, we trust that this approach will stimulate and lead to further investigations of design knowledge accumulation mechanisms.

7.2 Implications and Future Research

As practice-oriented DSR, our findings have implications for the research and practitioner communities: For the DSR community, we have added a novel perspective on reference models. While the research into conceptual modeling has strongly focused on modeling techniques and languages (Frank et al., 2014; Wand & Weber, 2002), we draw attention to the reference model as a design boundary object, as introduced by Bergman et al. (2007). Thus, our findings suggest reconsidering the roles of reference models, beyond the representation of generic design knowledge, their situational configurations, and their adaptation to create company-specific solutions

(Fettke & Loos, 2007; Thomas, 2006). As a design boundary object, a reference model serves as a vehicle for creating ontological foundations about sociotechnological phenomena and for accumulating and integrating knowledge from heterogeneous domains involving practitioner and academic communities. Thus, the ontological foundations will become the nucleus for studying specific practices and principles and for analyzing how capabilities are created and improved over time. We understand reference modeling as an opportunity to accumulate knowledge and create relevant DSR artifacts relating to wicked management problems. This seems particularly important in the context of digitalization, where the knowledge base that IS practice and research have developed in the past decades is often obsolete or simply ignored because it is not adapted and extended to the evolving problem space.

Our study also has implications for conducting DSR in a way that fosters knowledge accumulation and reuse. Based on our review of the data management field, we found that knowledge accumulation is often hampered by fragmented research activities and a lack of practitioner-researcher interaction. Typical DSR studies focus on artifact design, but seldom on reuse. Our case study provides insights into how longitudinal DSR processes can be organized in order to lead to relevant and theoretically grounded results: On the one hand, we found that institutionalized, multilateral industry-research collaborations foster knowledge accumulation. Compared to individual research projects, they provide a stable research context, create trust between practitioners and researchers, and enable coordinated research activities following DSR guidelines. On the other hand, suitable research topics and results depend on the maturity of the problem and solution spaces in a field, which evolves over time. As an important implication, researchers need to plan longitudinal DSR programs alongside the different knowledge accumulation stages, starting with the framing of problem and solution spaces (ontology), and moving toward detailed practices and implementation aspects (capability building), and possibly extending to reorientation. Ontological foundations and consistent documentation of artifacts enable the reuse of knowledge throughout the research process. The integration of specialized design knowledge into an overarching artifact (i.e., the reference model) allows for higher levels of empirical and theoretical knowledge aggregation.

For practitioners, our research implies that they can be consumers of reference models but also contributors to reference modeling. As consumers, they should opt for reference models with transparent and rigorous development processes that integrate practitioner and academic knowledge and that evolve over time. As contributors, they can benefit from sharing their knowledge, since reference models use the "wisdom of the many." Thus, the more contributors add knowledge in the design process, the more useful the resulting model will be.

7.3 Limitations

As with any retrospective analysis, our research design has limitations. Given the long period of research (12 years), we are unable to report on every aspect of reference model development and deliberately did not go into details about the different design areas. Although we have collected and analyzed very rich and comprehensive data, this paper can only present the decisive events during the research process and the main changes to the artifact. Although a retrospective analysis always comes with higher risks of biases (Leonard-Barton, 1990), it is an acknowledged field research method that provides unique insights into longitudinal processes (van de Ven & Huber, 1990) and thus seems to be very appropriate for analyzing knowledge accumulation in DSR. In our setting, the close interaction between researchers and firms in artifact design can be seen as a double-edged sword. It allows for unique access to field observations over more than a decade, but also creates certain risks, i.e., researchers reporting "only the processes of a disturbed system" (van de Ven & Huber, 1990, p. 216) instead of observing "the system in its natural state." The vast amount of raw data collected also creates barriers to the analysis and the communication of the research process (Leonard-Barton, 1990). To address these limitations and to reduce biases, we analyzed event histories and artifact evolution based on published sources and using very systematic analysis. To address the theoretical issues in longitudinal studies (Ployhart & Vandenberg, 2010), i.e., to conceptualize the form of change and to articulate the level of change, we developed a coding scheme and process to support the systematic analysis of changes in the design requirements and the artifact design over time.

Acknowledgments

This article summarizes insights from 12 years of consortium research at the Competence Center Corporate Data Quality (CC CDQ). The authors would like to thank all CC CDQ partner companies for their financial support, their passion for data management, and their active contributions to this unique industry-research collaboration!

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.

Aiken, P., Gillenson, M., Zhang, X., & Rafner, D. (2013). Data Management and Data administration: Assessing 25 years of practice. *Innovations in Database Design, Web Applications, and Information Systems Management*, *22*(3), 289-309.

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.

Amit, R., & Schoemaker, P. J. H. (1993). Strategic assets and organizational rent. *Strategic Management Journal*, *14*(1), 33-46.

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.

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. *Proceedings of the 47th Hawaii International Conference on System Sciences*.

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.

Bärenfänger, R., & Otto, B. (2015). Proposing a capability perspective on digital business models. *Proceedings of the 17th Conference on Business Informatics* (pp. 17-25).

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, *41*(3), 1-52.

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* (pp. 27-58). Physica-Verlag HD.

Bergman, M., Lyytinen, K., & Mark, G. (2007). Boundary objects in design: An ecological view of design artifacts. *Journal of the Association for Information Systems*, *8*(11), 546-568.

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.

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.

Buhl, H. U., Röglinger, M., Moser, F., & Heidemann, J. (2013). Big data. *Business & Information Systems Engineering*, *5*(2), 65-69.

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, 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.

Clarke, R. (2016). Big data, big risks. *Information Systems Journal*, *26*(1), 77-90.

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: Big data and the changing context of strategy. *Journal of Information Technology*, *30*(1), 44-57.

DalleMule, L., & Davenport, T. H. (2017). What's your data strategy? *Harvard Business Review*, *95*(3), 112-121.

DAMA. (2009). *The DAMA guide to the data management body of knowledge: (DAMA-DMBOK guide)* (First edition). Bradley Beach, NJ: Technics Publications LLC.

DAMA. (2017). *DAMA-DMBOK: Data management body of knowledge* (2nd edition). Technics Publications.

EDM Council. (2018). *Data capability assessment model (dcam) overview*. https://cdn.ymaws.com/edmcouncil.org/resource/resmgr/featured_documents/EDMC_DCAM_Overview.pdf

EFQM. (2009). *EFQM excellence model*. *EFQM model 2010*. EFQM.

EFQM. (2011). *EFQM framework for corporate data quality management: Assessing the* *organization's data quality management capabilities*. EFQM.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: opportunities and challenges. *The Academy of Management Journal*, *50*(1), 25-32.

Feldman, M. S., & Orlikowski, W. J. (2011). Theorizing practice and practicing theory. *Organization Science*, *22*(5), 1240-1253.

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. (2007). Perspectives on reference modeling. In P. Fettke & P. Loos (Eds.), *Reference Modeling for Business Systems Analysis* (pp. 1-21). IGI Global.

Frank, U. (2007). Evaluation of reference models. In P. Fettke & P. Loos (Eds.), *Reference modeling for business systems analysis* (pp. 118-140). IGI Global.

Frank, U., Strecker, S., Fettke, P., vom Brocke, J., Becker, J., & Sinz, E. (2014). The Research field "modeling business information systems." *Business & Information Systems Engineering*, *6*(1), 39-43.

Gartner. (2014). *The Gartner enterprise information management framework.* https://blogs.gartner.com/andrew_white/files/2016/10/On_site_poster.pdf

Gerring, J. (2006). *Case study research: Principles and practices*. Cambridge University Press.

Gillenson, M. L. (1985). Trends in data administration. *MIS Quarterly*, *9*(4), 317-325.

Goldkuhl, G. (2011). The research practice of practice research: Theorizing and situational inquiry. *Systems, Signs & Actions, 5*(1), 7-29.

Goldkuhl, G. (2012). From action research to practice research. *Australasian Journal of Information Systems*, *17*(2), 57-78.

Goldstein, R. C., & McCririck, I. B. (1981). The Stage hypothesis and data administration: Some contradictory evidence. *Proceedings of the International Conference on Information Systems* (pp. 309-324).

Goodhue, D. L., Kirsch, L. J., Quillard, J. A., & Wybo, M. D. (1992). Strategic data planning: Lessons from the field. *MIS Quarterly*, *16*(1), 11-34.

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.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337-355.

Gregor, S., Müller, O., & Seidel, S. (2013). Reflection, abstraction and theorizing in design and development research. *Proceedings of the European Conference on Information Systems.*

Grover, V., & Teng, J. T.C. (1991). How effective is data resource management? Reassessing strategic objectives. *Journal of Information Systems Management*, *8*(3), 16-20.

GS1. (2010). *Data Quality Framework: Version 3.0*. Issue 2. https://www.gs1.org/services/data-quality/data-quality-framework

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, *53*(8), 1049-1064.

Haug, A., & Stentoft Arlbjørn, J. (2011). Barriers to master data quality. *Journal of Enterprise Information Management*, *24*(3), 288-303.

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.

Hopkins, B., Leganza, G., Goetz, M., & Lee, J. (2018). *Stretch your data management capabilities: continuous improvement: The data management playbook* (No. 87742). https://reprints.forrester.com/#/assets/2/220/RES87742/reports

Hüner, K. M., Ofner, M. H., & Otto, B. (2009). Towards a maturity model for corporate data quality management. *Proceedings of the ACM Symposium on Applied Computing.*

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.* IBM Corporation.

International Organization for Standardization (ISO). (2011). *Data quality—Part 150: Master data: Quality management framework: ISO 8000-150:2011.*

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.

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, *53*(1), 148.

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.

Leonard-Barton, D. (1990). A dual methodology for case studies: synergistic use of a longitudinal single site with replicated multiple sites. *Organization Science*, *1*(3), 248-266.

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.

Marino, K. E. (1996). Developing consensus on firm competencies and capabilities. *Academy of Management Perspectives*, *10*(3), 40-51.

Marsh, R. (2005). Drowning in dirty data? It's time to sink or swim: A four-stage methodology for total data quality management. *Journal of Database Marketing & Customer Strategy Management*, *12*(2), 105-112.

Mathiassen, L. (2002). Collaborative practice research. *Information Technology & People*, *15*(4), 321-345.

Meuser, M., & Nagel, U. (2009). The expert interview and changes in knowledge production. In A. Bogner, B. Littig, & W. Menz (Eds.), *Interviewing experts* (pp. 17-42). Palgrave Macmillan UK.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE.

Möller, K., Otto, B., & Zechmann, A. (2017). Nutzungsbasierte Datenbewertung. *Controlling*, *29*(5), 57-66.

Mumford, E. (2006). The story of socio-technical design: Reflections on its successes, failures and potential. *Information Systems Journal*. (16), 317-342.

Niederman, F., & March, S. T. (2012). Design science and the accumulation of knowledge in the information systems discipline. *ACM Transactions on Management Information Systems*, *3*(1), 1-15.

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.

Orr, K. (1998). Data quality and systems theory. *Communications of the ACM*, *41*(2), 66-71.

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, pp. 3-18). Springer.

Otto, B. (2011a). A morphology of the organisation of data governance. *Proceedings of the European Conference on Information Systems (ECIS 2011)*.

Otto, B. (2011b). Quality management of corporate data assets. In C.-P. Praeg & D. Spath (Eds.), *Quality management for IT services* (pp. 193-209). IGI Global.

Otto, B. (2012a). *Enterprise-wide data quality management in multinational corporations* (Cumulative postdoctoral thesis, University of St. Gallen, St. Gallen, Switzerland).

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. *Proceedings of the 16th Americas Conference on Information Systems.*

Otto, B., Hüner, K. M., & Österle, H. (2009). Identification of business oriented data quality metrics. *Proceedings of the 14th International Conference on Information Quality* (pp. 122-134).

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., & Ofner, M. H. (2010). Towards a process reference model for information supply chain management. *Proceedings of the European Conference on Information Systems*.

Otto, B., & Reichert, A. (2010). Organizing master data management: findings from an expert survey. *Proceedings of the 2010 ACM Symposium on Applied Computing* (pp. 106-110).

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., Legner, C., & Otto, B. (2017). Towards a reference model for data management in the digital economy. *Proceedings of the 12th International Conference on Design Science Research in Information Systems and Technology* (73-82).

PIC. (2016). *Data quality maturity model: PIC Data quality working group guide.* https://www.pic.gov/sites/default/files/Data%20Quality%20Maturity%20Model%20and%20Example%20Practices%20Guidebook%20-%202016.pdf

Pinsonneault, A., & Kraemer, K. (1993). Survey research methodology in management information systems: An assessment. *Journal of Management Information Systems*, *10*(2), 75-105.

Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal research: the theory, design, and analysis of change. *Journal of Management*, *36*(1), 94-120.

Porter, M. E., & Heppelmann, J. E. (2015). How smart, connected products are transforming companies. *Harvard Business Review*, *93*(10), 96-114.

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.

Ravindra, P. S. (1986). Data administration: an old function adopts to its new role. *Journal of Information Systems Management*, *3*(4), 47-51.

Recker, J., Rosemann, M., van der Aalst, W. M.P., Jansen-Vullers, M., & Dreiling, A. (2007). Configurable reference modeling languages. In P. Fettke & P. Loos (Eds.), *Reference Modeling for Business Systems Analysis* (pp. 22-46). IGI Global.

Redman, T. C. (1995). Improve data quality for competitive advantage. *Sloan Management Review*, *36*(2), 99-107.

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

Reichert, A., Otto, B., & Österle, H. (2013). A Reference process model for master data management. *Proceedings of the 11th International Conference on Wirtschaftsinformatik* (pp. 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.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155-169.

Schemm, J. W. (2008). Stammdatenmanagement zwischen Handel und Konsumgüterindustrie: Referenzarchitektur für die überbetriebliche Datensynchronisation (Doctoral dissertation, University of St. Gallen, St. Gallen, Switzerland).

Schermann, M., Böhmann, T., & Krcmar, H. (2009). Explicating design theories with conceptual models: Towards a theoretical role of reference models. In J. Becker, H. Krcmar, & B. Niehaves (Eds.), *Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik* (pp. 175-194). Physica-Verlag HD.

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*.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS Quarterly*, *35*(1), 37-56.

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.

Proceedings of the 8th International Conference on Information Quality (pp. 344–352).

Sheng, Y. P., & Mykytyn, P. P. (2002). Information technology investment and firm performance: a perspective of data quality. *Proceedings of the 7th International Conference on Information Quality* (pp. 132-141).

Spruit, M., & Pietzka, K. (2015). MD3M: The master data management maturity model. *Computers in Human Behavior*, *51*, 1068-1076.

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.

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.

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.

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, pp. 484-496). Springer.

Thomas, O. (2007). Version management for reference models: design and implementation. In J. Becker & P. Delfmann (Eds.), *Reference Modeling* (pp. 1-26). Physica-Verlag HD.

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, pp. 121-143). Springer US.

Van de Ven, A. H., & Huber, G. P. (1990). Longitudinal field research methods for studying processes of organizational change. *Organization Science*, *1*(3), 213-219.

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? *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*.

vom Brocke, J. (2007). Design principles for reference modeling. In P. Fettke & P. Loos (Eds.), *Reference modeling for business systems analysis* (pp. 47-76). IGI Global.

vom Brocke, J., & Buddendick, C. (2006). Reusable conceptual models: Requirements based on the design science research paradigm. *Proceedings of the First International Conference on Design Science Research in Information Systems and Technology* (pp. 576-604).

Wand, Y., & Weber, R. (2002). Research commentary: information systems and conceptual modeling—A research agenda. *Information Systems Research*, *13*(4), 363-376.

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.

Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all: A contingency approach to data governance. *Journal of Data and Information Quality*, *1*(1), 1-27.

Weldon, J.-L. (1981). *Data base administration*: *Applications of modern technology in business*. Springer.

Wende, K. (2007). A model for data governance: organising accountabilities for data quality management. *Proceedings of the 18th Australasian Conference on Information Systems* (pp. 417-425).

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.

Xie, K., Wu, Y., Xiao, J., & Hu, Q. (2016). Value co-creation between firms and customers: The role of big data-based cooperative assets. *Information & Management*, *53*(8), 1034-1048.

Yin, R. K. (2014). *Case study research: Design and methods* (5th edition). SAGE.

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.

Appendix

Table A1. Reference Models for Data Management

Contribution	Publications	Origin	Description
Master data management maturity model	(Spruit & Pietzka, 2015)	Research	A maturity model that focuses on master data with five key topics: (1) data model, (2) data quality, (3) usage and ownership, (4) data protection, and (5) maintenance. Each key topic has a number of focus areas (13 in total), which can be described on five maturity levels, resulting in 65 capability stages.
DAMA-DMBOK framework	(DAMA, 2009, DAMA, 2017)	Industry consortium	Reference model, which has evolved through several versions since its first publication in 2006. The current version—called the DAMA-DMBOK2 Data Management Framework (The DAMA Wheel)—has 11 knowledge areas: (1) data governance, (2) data modeling and design, (3) data storage and operations, (4) data security, (5) data integration and interoperability, (6) document and content management, (7) reference and master data, (8) data warehousing and business intelligence, (9) metadata, (10) data duality, and (11) data architecture. Each knowledge area comes with a list of activities defined for it (adding up to 102 activities). Each activity needs to be executed in four phases: (1) plan, (2) develop, (3) control, and (4) operate.
Data quality maturity model	(PIC, 2016)	Industry consortium	A maturity model with four elements: (1) policies and procedures, (2) quality control and assurance practices, (3) governance and leadership (including culture), and (4) human capital. For each element, the model specifies roughly four maturity stages and provides brief recommendations for proceeding to the next maturity stage.
Data capability assessment model	(EDM Council, 2018)	Industry consortium	A maturity model with eight components: (1) data management strategy, (2) data management business case, (3) data management program, (4) data governance, (5) data architecture, (6) technology architecture, (7) data quality, and (8) data control environment. These components are further subdivided into 36 capabilities, 112 subcapabilities, and 306 objectives.
Data governance maturity model	(IBM Data Governance Council, 2007)	Industry consortium	A maturity model with four groups: (1) enablers, (2) outcomes, (3) core disciplines, and (3) supporting disciplines, each comprising a number of data governance domains (11 in total).
Data quality management system	(GS1, 2010)	Standardization body	Structured as a matrix that comprises four functional areas in the horizontal direction: (1) organizational capabilities, (2) policies and standards, (3) business processes, and (4) systems capabilities; and four main activity types in the vertical direction: (1) plan, (2) document, (3) execute, and (4) monitor. For each field of the matrix, several capabilities are defined (73 in total).
Master data quality management framework	(ISO, 2011)	Standardization body	Defines three top-level processes: (1) data operations, (2) data quality monitoring, and (3) data quality improvement; each has three lower-level processes.
Data management capability model	(Hopkins et al., 2018)	Analyst	A reference model that defines three value streams: (1) data management planning and data architecture development, (2) service delivery, and (3) security and governance; each has a number of capabilities (9 in total)
Enterprise information management maturity model	(Gartner, 2014)	Analyst	A maturity model that has evolved since its first publication in 2008. It has seven building blocks: (1) vision, (2) strategy, (3) metrics, (4) information governance, (5) organization and roles, (6) information lifecycle, and (7) information infrastructure.

Big data analytics capability model	(Akter et al., 2016)	Research	A hierarchical reference model with three primary capability dimensions (management, technology, and talent) and 11 subdimensions.
Big data resources framework	(Gupta & George, 2016)	Research	A hierarchical reference model including three classes of resources (i.e., tangible, human, and intangible) and seven resources which, combined with one another, allow one to build a big data analytics capability. For each resource, between two and six items (32 in total) are presented, which specify the requirements to be met.

Table A2. Decisive Events in the Research Process

ID	Date	Technique	DSR phase	Decisive event
1	23.11.2006	Focus group	Problem formulation (P), solution design (D)	The identification of requirements and motivation for data quality management (i.e., business alignment, compliance, M&A). Discussion of a first reference model draft
2	07.03.2007	Expert interview	P	The identification of challenges to data management (i.e., alignment with business processes, organizational setup in regions, missing top management support) Understanding of data as a product with a price Need for a cockpit for data quality.
3	25.06.2007	Plenary discussion	D	An ontological discussion of data management terms An understanding of data management as an infrastructure Data types define the data management's scope
4	25.06.2007	Plenary discussion	D	A discussion of existing data management reference models: while the contents of existing solutions are mainly appropriate, their structure is inconsistent, resulting in the need for a consistent model
5	15.11.2007	Focus group	D	Consensus building about the alpha version of the reference model (the CDQM reference model)
6	08.01.2008	Expert interview	P	The identification of challenges to data management (i.e., defining the organizational setup and the data management's scope, measuring the data quality management's maturity to steer the activities)
7	28.01.2008	Expert interview	Instantiation (I), P	A company-specific instantiation of the CDQM reference model, including a maturity model, indicated the need for a generic maturity assessment for CDQM
8	11.04.2008	Expert interview	P	The identification of challenges to data management (i.e., the demand for continuous resources for data management, identified as data management, is understood as a permanent capability and not a one-off project, continuous improvement as a relevant aspect of data management, and data management as a capability)
9	16.09.2008	Project	P, I	A three-day education program for 20 employees about the data management based on the CDQM reference model, underlining that education is a relevant aspect for capability building
10	05.02.2009	Expert interview	P	The identification of challenges to data management (i.e., despite structured CDQM approach problems to position the topic at the executive level, data management is multidimensional, and needs time to establish, roles involved at central and local levels, multiple functions, processes, and units are involved, various data domains are affected and in scope)
11	17.02.2009	Plenary discussion	P, D	The need to determine the value of (product) data management—a cost approach (based on the common cost accounting method) was applied
12	13.10.2009	Plenary discussion	P	A discussion of data management's boundaries (i.e., a focus on product and on customer data)

				A discussion of organizational design options (i.e., a need for a central product data management organization, but also for local responsibilities)
13	26.11.2009	Case study	D, I	An opportunity cost-based method for defining (product) data management's business value
14	01.12.2009	Plenary discussion	D	Service orientation as a design principle for data management
15	13.04.2010	Focus group	D	An analysis of existing capability and maturity models to be used for quality-oriented data management
16	09.09.2010	Project	I	The first assessment using the maturity model
17	23.09.2010	Focus group	D	A discussion of the understanding of data management (i.e., data as a product, "data management factories," service orientation of data management)
18	25.11.2010	Plenary discussion	P	Expansion of the maturity model toward a tool for benchmarking data management activities
19	02.02.2011	Plenary discussion	P	A discussion of business scenarios for "master data management 2015" confirmed the original rationale
20	02.02.2011	Plenary discussion	D	The maturity approach and benchmarking concepts were confirmed; however, the latter never really took off
21	06.04.2011	Project	I	A maturity model design was demonstrated and proved to be applicable
22	29.11.2011	Expert interview	D	Establish data management in stages. Data management is a capability, not a function
23	12.02.2012	Focus group	P	The need to enhance data management's scope to include consumer data An alpha version of the reference model—originally designed for product master data—was also applicable to further data domains; however, the terminology used did not relate to marketing and sales representatives
24	06.03.2012	Focus group	P	An unsuccessful attempt to establish a further, separate research consortium on consumer-centric information management. Consumer-centricity as a topic anchored in marketing and sales could not be connected to the ontological basis of data management
25	18.04.2012	Plenary discussion	P	The need to integrate external data (i.e., from smart meters and e-mobility) and to fulfill compliance requirements in data management
26	22.06.2012	Focus group	P	The need to assess the influences of new technologies such as big data analytics on data management and to expand the scope toward data security management
27	13.02.2013	Plenary discussion	P	The need for a profitability analysis of data management
28	09.10.2013	Focus group	P	The need for data architectures for big data scenarios
29	14.11.2013	Project	I	Elaboration of detailed material to support an executive education program for data management based on the alpha version of the reference model
30	29.10.2014	Focus group	P	The need to address the requirements of digitalization and Industry 4.0 in data management
31	10.12.2014	Focus group	D	A revision of the maturity model
32	04.11.2015	Focus group	P	The need for a revised version of the reference model formulated. The need to develop data management services and capabilities for the digital economy
33	25.02.2016	Plenary discussion	P, D	A collection of requirements for a strategic data management reference model (i.e., the business criticality of data, the business value of data, data compliance and data security, coverage of all data types)

				A discussion of the continuous management cycle as the basic structure for the beta version of the reference model
34	28.04.2016	Focus group	P, D	The requirements of data management and for a revised reference model (i.e., business criticality of data, business value of data management) A review of existing reference and maturity models for data management A discussion of relevant design areas for data management (i.e., data lifecycle as a separate design area and not part of processes, change management is not part of data management reference model on the highest level, data strategy as part of the goals)
35	24.09.2016	Focus group	D	A discussion of the terminology for the design areas of the model's beta version (i.e., performance management instead of controlling, data excellence as an umbrella term for the internal results of data management)
36	10.11.2016	Focus group	P	The identification of requirements for data management (i.e., compliance with regulations such as GDPR, management of open data)
37	08.12.2016	Focus group, survey	D	A discussion of the terminology for the design areas of the model's beta version (i.e., people, roles, and responsibilities instead of organization, business value instead of business impact). Formative evaluation
38	20.02.2017	Project	I	A beta version of the reference model applied for communicating with stakeholders of data management and for developing a data strategy
39	22.02.2017	Focus group	D	A discussion of layout options and a decision on a graphic representation of the model
40	19.04.2017	Project	I	The instantiation of the reference model beta version for sensor data management proving the validity of the reference model for new data domains
41	04.09.2017	Expert interview	P, D	Requirements for the revision of the maturity model (i.e., compatibility with the previous maturity model, number of questions, reflection on the reference model beta version) Design of a draft version of the maturity model based on the reference model's beta version
42	05.10.2017	Project	I	An instantiation of the maturity model proving its utility
43	07.12.2017	Focus group	D	A refinement of maturity model elements and questions
44	21.02.2018	Focus group	P	Differences between managing big data in contrast to master and internal data (i.e., additional roles such as data scientists, processes such as data science usage case identification, and architectures for data lakes)
45	27.04.2018	Case study	P, D	An understanding that data management in digital, data-driven enterprises requires a dual approach (data foundation and data science)

Table A3. Design Areas of the Data Excellence Model (for Figure A1: Metamodel)

Design area (DA)	Description	Result documents	Practices and justificatory knowledge	Situational inquiry and empirical evidence
DA1: Business capabilities	Seeks to define a set of skills, routines, and resources a company needs in order to achieve business objectives.	• Business drivers and goals • Business capabilities map	• Data is a strategic resource (Goodhue et al., 1988). • Data is monetized in business models or value propositions, business processes, and decision-making (Schüritz et al., 2017).	• 1 case study • 1 focus group • (Bärenfänger & Otto, 2015)
DA2: Data management capabilities	Seeks to define a set of skills, routines, and resources a company needs in order to support business capabilities via data management.	• Data management capabilities map • Portfolio of data products and data management services	• Data management is a dynamic capability to deploy data resources (Otto, 2012a). • Data management is contingent on business objectives and capabilities (Jain et al., 1998)	• 7 focus groups
DA3: Data strategy	Seeks to evaluate a set of strategic choices around data management in order to be able to make decisions concerning the ways data are to be managed and used.	• Data management vision, objectives, and scope • Data management roadmap • Resource plan	• Data strategy defines the guiding policy for managing data (DalleMule & Davenport, 2017). • Data strategy is linked to corporate strategy and has mutual dependencies to function and divisional strategies.	• 6 case studies • 6 focus groups
DA4: People, roles, and responsibilities	Seeks to define skills and design and implement the organization and roles, to ensure effective data management and the consistent use of data across the entire organization.	• Role descriptions • Interaction model	• Decision rights and roles must be assigned so as to achieve consistent, company-wide data usage behavior. • Data is only managed if data ownership and data stewardship are trained and executed (Khatri & Brown, 2010).	• 3 case studies • 4 focus groups • (Weber et al., 2009) • (Otto & Reichert, 2010) • (Otto, 2011a)
DA5: Processes and methods	Seeks to define procedures and standards for proper and consistent data management and usage.	• Data management processes • Data management methods • Business rules	• Data management as a capability is implemented in organizational routines (Marino, 1996). • Methods assure standardized, enterprise-wide behavior in data management and data use (Khatri & Brown, 2010).	• 4 case studies • 3 focus groups • (Reichert et al., 2013)
DA6: Data lifecycle	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	• Core business and data objects • Data lifecycle processes • Data sources • Data consumers	• In line with the management of physical resources, managing the data lifecycle assures an enterprise-wide consistent approach to create, maintain, use, and archive data (Wang, 1998; Wang, Lee, Pipino, & Strong, 1998).	• 2 focus groups • (Otto & Ofner, 2010) • (Ofner et al., 2013)

	consumers, and data usage contexts.		• Orchestrate the data value chain for big data management (Abbasi et al., 2016)	
DA7: Data applications	Seeks to plan, implement, and maintain applications designed to create, maintain, use, and archive data and to ensure data excellence.	• Application • Functionalities • Interfaces • Storage	• Data applications provide the required functionalities for managing data. • The interfaces and storage of applications need to be documented and monitored to streamline data flows (Akter et al., 2016; Bourdreau & Couillard, 1999; Sun et al., 2006).	• 1 case study • 4 focus groups • (Otto et al., 2012)
DA8: Data architecture	Seeks to define and maintain specifications that provide a shared business vocabulary, express strategic data requirements, and outline high-level integrated application system landscape designs and data flows (for storing and distributing data of enterprise-wide validity).	• Data models • Data storage and distribution architecture • Data flow	• For core business objects and their attributes, both the leading applications for storage and distributions and the consuming applications and the interfaces need to be documented. • Core data entities and their relationships are described by data models (Brancheau, Schuster, & March, 1989).	• 3 case studies • 4 focus groups • (Baghi et al., 2014)
DA9: Performance management	Seeks to plan, implement, and control all activities for measuring, assessing, improving, and ensuring data excellence as well as the performance of data management as an organizational capability.	• Performance management system • Data excellence metrics • Business value metrics	• A performance management system supports enterprises, conveying their goals through analyzing, measuring, and controlling the progress and outcomes of data management (Ferreira & Otley, 2009).	• 4 case studies • 5 focus groups • (Otto et al., 2009)
DA10: Data excellence	Refers to data management's impacts on the data, first concerning data quality (defined as *fitness for purpose*), but also concerning additional data-related aspects, such as data compliance, data security, and data privacy.	• Data excellence (dimensions: data quality, data compliance, data security, data privacy)	• Creating transparency and communicating progress and performance is the basis for continuous improvement (Batini et al., 2009; EFQM, 2009). • Data excellence (Suarez, Calvo-Mora, & Roldán, 2016) comprises the traditional goals of providing high-quality data (Batini & Scannapieca, 2006; English, 2003; Wang, 1998; Wang et al., 1998) and addresses data compliance, data security, and data privacy (Delbaere & Ferreira, 2007; Sadeghi, Wachsmann, & Waidner, 2015).	• 1 case study • (Hüner et al., 2011)

DA11: Business value	Refers to data management's impacts on business concerning financials, business processes, customers, and organizational growth.	• Business value	• Data excellence impacts on business performance (Joshi & Rai, 2000; Sheng & Mykytyn, 2002; Spruit & Pietzka, 2015). • Creating transparency and communicating the value to the business generated by data management improves the acceptance of data management in the enterprise (Chen et al., 2012; LaValle et al., 2011).	• 1 case study • 5 focus groups • (Otto, 2012b) • (Möller, Otto, & Zechmann, 2017)

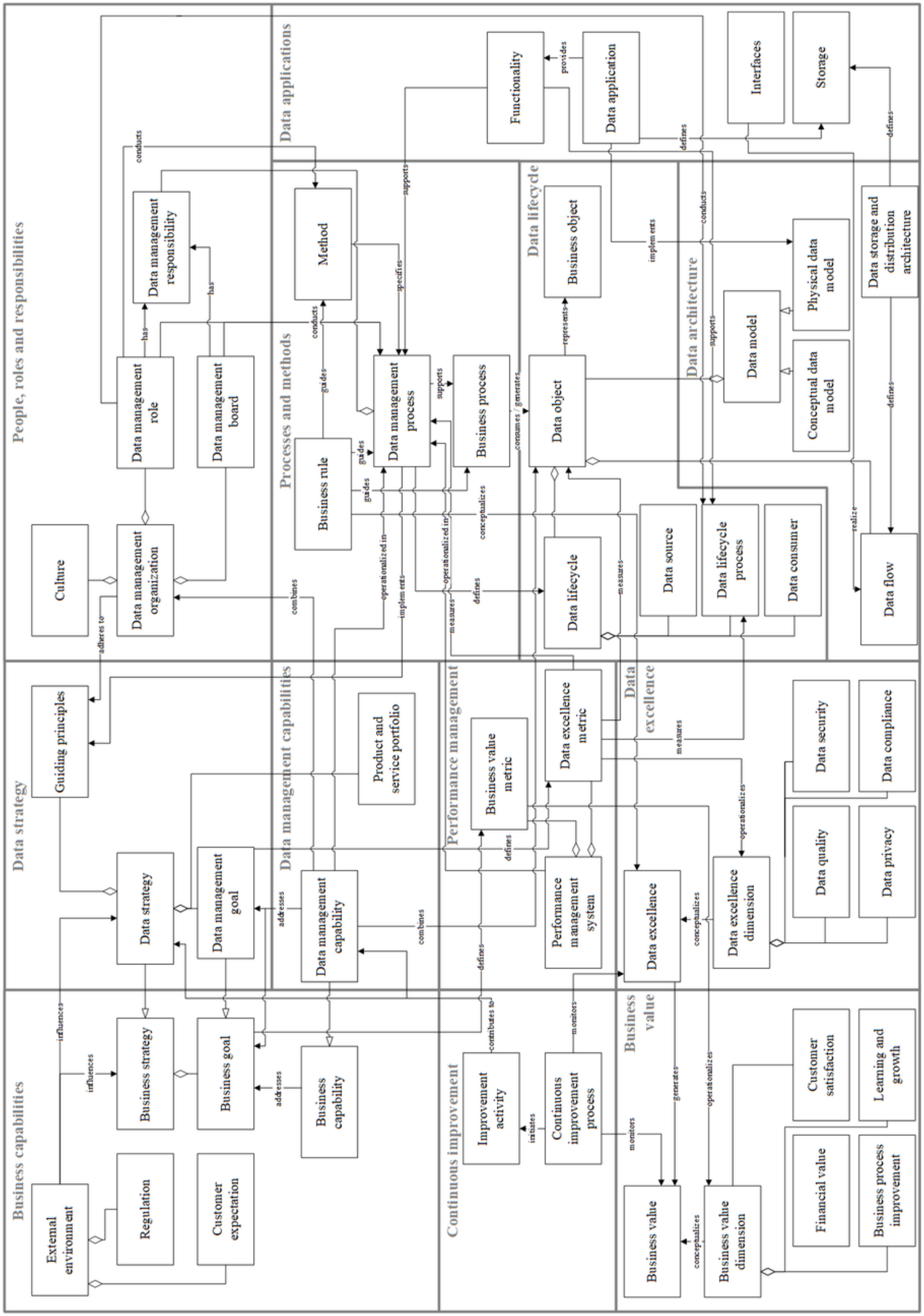

Figure A1. Metamodel

Table A4. Comparison of the Suggested Reference Model with Other Data Management Frameworks

	Design process	Knowledge accumulation	Rationale: Compliance	Rationale: Data privacy	Rationale: Data security	Rationale: Data quality	Boundaries: Master data	Boundaries: Big data	Design area: Business capabilities	Design area: Data management capabilities	Design area: Data strategy	Design area: People, roles, and responsibilities	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
Alpha version (the CDQM reference model)	Consortium research (multilateral, longi-tudinal DSR process)	no	-	-	-	X	X	-	-	-	X	X	X	(x)	X	X	X	-	-	(x)
Beta version (the Data Excellence Model)	Consortium research (multilateral, longi-tudinal DSR process)	Alpha version: 2007/11 (see above) Beta version: 2017	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Master data management maturity model	Design science research	no	-	(x)	X	X	X	-	-	-	-	X	-	X	-	X	X	(x)	X	X
DAMA-DMBOK framework	Consensus building in consortium *(details not reported)*	First version: 2006 Intermediate versions between 2006 and 2017 Latest version: 2017	X	X	X	X	X	(x)	(x)	(x)	X	X	X	X	X	X	X	(x)	X	X
Data quality maturity model	9 focus groups, survey	no	X	-	-	X	X	-	-	-	-	X	X	-	-	-	X	-	-	(x)
Data capability assessment model	Consensus building in consortium *(details not reported)*	no	X	X	X	X	X	-	-	-	X	X	X	X	X	X	X	(x)	X	-
Data governance maturity model	Consensus building in consortium *(details not reported)*	no	X	X	X	X	X	(x)	X	X	-	X	X	X	(x)	X	X	(x)	X	-
Data quality management system	*Not reported*	no	(x)	(x)	X	X	X	X	-	-	X	X	X	X	X	(x)	X	X	(x)	X
Master data quality management framework	Consensus building in standardization body *(details not reported)*	no	-	-	-	X	X	X	-	-	(x)	X	X	X	(x)	X	X	(x)	-	X
Data management capability model	*Not reported*	no	-	-	X	-	X	(x)	(x)	X	-	(x)	X	(x)	X	X	-	-	-	-
Enterprise information management maturity model	*Not reported*	First version: 2008 Latest version: 2014	-	-	-	X	X	X	(x)	(x)	X	X	(x)	X	X	-	X	(x)	X	X
Big data analytics capability model	Literature review, Delphi studies	no	-	-	-	(x)	X	X	(x)	X	X	X	-	-	X	(x)	X	-	-	-
Big data resources framework	Scale development procedure	no	-	-	-	-	X	X	-	-	(x)	X	-	-	X	-	-	-	-	(x)

X: fully addressed; (x): partially addressed; -: not addressed

This review of competing artifacts includes only artifacts developed by researchers, industry consortia, analysts, and standardization bodies. We have excluded reference models from consulting firms or software vendors, since these models tend to be single-expert and/or single-case induced. Further, we considered only reference or maturity models for data management as relevant state-of-the-art, and that were publicly available and sufficiently detailed.

About the Authors

Christine Legner is a professor of information systems at HEC Lausanne—the Faculty of Business and Economics at the University of Lausanne, Switzerland. She also serves as the academic director of the Competence Center Corporate Data Quality (CC CDQ), an industry-funded research consortium and expert community in the data management field. She received her PhD and postdoctoral qualification from the University of St. Gallen. Her research interests relate to data management and data quality, business information systems, strategic IT planning, and enterprise architecture.

Tobias Pentek is the head of Community and Innovation at the Competence Center Corporate Data Quality (CC CDQ). He has broad experience in (data) management consulting with multiple projects across industries, developing data strategies and data governance. He is a doctoral candidate at the University of St. Gallen. He received a bachelor's degree in general management from the EBS Business School and a master's in business management from the University of St. Gallen. His research interests relate to data strategies and data management methods.

Boris Otto is a professor of industrial information management at TU Dortmund University and the executive director of the Fraunhofer Institute for Software and Systems Engineering (ISST) in Dortmund, Germany. He holds a doctoral degree in mechanical engineering from the University of Stuttgart and obtained his postdoctoral qualification from the University of St. Gallen. His research focuses on data resource management, data quality management, data sharing, and data-driven business models.