Towards a Reference Model for Data Management in the Digital Economy

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

[1] University of St. Gallen, Switzerland
tobias.pentek@student.unisg.ch
[2] University of Lausanne, Faculty of Business and Economics (HEC), Switzerland
christine.legner@unil.ch
[3] TU Dortmund University, Germany
boris.otto@tu-dortmund.de

Abstract. The digital and data-driven economy requires enterprises from all industries to revisit their existing data management approaches. To address the changing and broader scope of data management activities in the digital economy, this research in progress paper proposes a reference model, that describes the design areas of data management.

Keywords: Data management · Reference model · Data governance · Data architecture · Master data · Big data

1 Introduction

In recent years, the digitalization of processes and the digital transformation of business models has altered many industries, and data is regarded as a key resource for companies enabling new products and services [1]. For example, Industry 4.0 scenarios in the manufacturing industry are transforming logistics and production processes. In the future, a smart product will steer itself through the production process, order transportation services, and inform the manufacturing stations about the required assembly steps [2]. For fulfilling this vision, data from multiple sources and sensors need to be gathered, aggregated, maintained, enriched, and provided.

Understanding data as a strategic and valuable resource is a change for corporations and requires them to broaden the scope, challenge existing practices, and refine current approaches of data management. Traditionally, data management comprises "policies, practices and projects that acquire, control, protect, deliver and enhance the value of data and information" [3]. To guide practitioners in the implementation and conduction of data management, a number of data management frameworks and reference models has been suggested. However, these have been designed in the 1990s and 2000s, and it is unclear whether they address the requirements of the digital and data-driven economy.

Following the design research paradigm, this paper identifies requirements and proposes a reference model for effective data management in the digital and data-driven economy. In accordance with Fettke and Loos [4], this reference model describes solution patterns for data management professionals and researchers and provides recommendations for conducting effective data management. It has been developed based on a systematic action design research (ADR) approach in a joint effort with more than 15 European companies and researchers from three universities. The paper reports on the structure and design areas of the reference model as well as on the first evaluation results.

2 Background

Data Management

Data management aims at the efficient usage of data in companies [5]. Academic literature has traditionally elaborated on the role and importance of data assets as well as on data quality. It comprises all management tasks of the data lifecycle on a strategic, governing, and technical level [6]. Data management includes the formulation of a data strategy [7], the definition of data management processes, standards, and measures, the assignment of roles and responsibilities [8], the description of the data lifecycle and architecture – covering data models and data modeling standards – [9], and the management of applications and systems [10].

To guide practitioners in the implementation of data management, various data management frameworks and reference models have been suggested. These include contributions from consortia – such as the DAMA-DMBOK Functional Framework [11], the CDQ Framework [12], and the DGI Data Governance Framework [13] –, from consulting firms – including Capgemini [14], Gartner [15], Forrester [16], and Infosys [17] –, from software vendors – like Oracle [18], IBM [19], Informatica [20], and SAS [21] – , and with GS1 [22] from a standardization body. The majority of these frameworks as well as a large number of research considers the provision of high quality data as the most important goal of data management [23].

Research Gap: Data Management in the Digital Economy

Data is the key resource of companies in the digital economy. The term “digital economy” considers the digitalization of processes and the digital transformation of business models in corporations as well as the digitalization of the society [24]. It covers various trends such as Industrie 4.0, the Internet of Things (IoT), and Big Data and is mainly driven by technological advances, which led to falling prices for sensors and cheaper and faster processing of data. While data is considered a key concern in digitalization initiatives, the advances of data management are still limited. The majority of existing data management frameworks has a narrow scope and focuses on the quality of master data from company-internal sources. However, with the growing importance of data in the digital economy, data managers are challenged by additional requirements to fulfill compliance, privacy, and security concerns in their activities and to include further data types such as streaming data or data from external sources. There are first publications to address these aspects, such as data management for big data [25–27]. Nevertheless, to the best of our knowledge, a comprehensive reference model providing guidance to data managers and researchers in the context of the digital economy does not exist. Our research addresses this gap by answering the following research question: How to design data management for the digital and data-driven economy?

3 Research Approach

Our research aims at developing a reference model that outlines the main design areas to be addressed by companies to effectively manage data as corporate resource. A reference model specifies the generally valid elements of a system that can serve as a reference for designing company-specific models [4]. Our research objectives thereby consist of developing prescriptive knowledge as described by Gregor's [28] type V theory and constructing information system-related problem solutions. The emerging artefact has been designed in a consortium research program [29] since February 2016, following ADR. ADR combines design science research (DSR) and action research (AR) and constitutes "a research method for generating prescriptive design knowledge through building and evaluating ensemble IT artifacts in an organizational setting" [30]. Senior data management professionals from more than 15 European enterprises in various industries and researchers from three academic institutions have contributed to the artifact over a period of more than twelve months. After an initial discussion about the need for a reference model at a consortium workshop in February 2016, the requirements and the emerging reference model were discussed and evaluated at five consortium workshops between April 2016 and February 2017. **Fig. 1** provides an overview of the four main stages of ADR, which are described in the following.

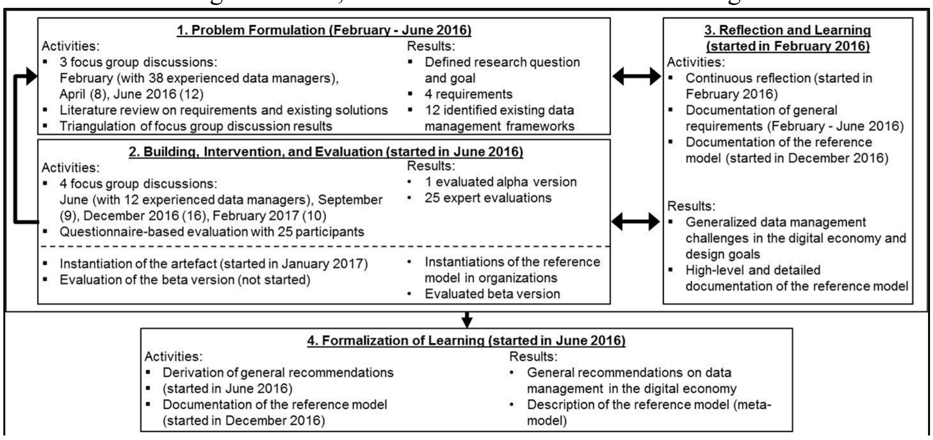

Fig. 1. Adopted ADR Process on the Basis of Sein et al. [30]

Problem formulation: The research activities were initiated by experienced data managers in the consortium research program. Although these data managers were using established data management frameworks, they were lacking guidance for facing the new data-related challenges in the digital economy. Based on this call for action, the specific requirements as well as existing solutions were discussed during consortium workshops in three focus groups in February (with 38 experienced data managers), April (8), and June 2016 (12). The discussion results were documented and triangulated with scientific publications by the researchers.

Building, intervention, and evaluation (BIE) includes two cycles. As suggested by Sein et al. [30], the first BIE cycle aims at developing an early alpha version of the artefact, which is a stable version that will be instantiated, repeatedly tested, and continuously refined in the second cycle. In the first cycle, the emerging artefact was repeatedly discussed and refined by practitioners in four focus groups during consortium workshops in June (with twelve experienced data managers), September (9), December 2016 (16), and February 2017 (10). Each of the four sessions lasted two hours, was moderated by the same researcher and observed by a second researcher. The discussion results were documented and formed the basis for further adjustments of the artefact. The first three focus group sessions focused on the structure, design areas, content, and naming of the reference model, while the last session concentrated on the graphical visualization. Section 4 outlines the resulting alpha version of the artefact. In addition to evaluations in focus groups discussions, we conducted a questionnaire-based evaluation in December 2016. The results are presented in section 5. In the second BIE cycle, we are instantiating the artifact in selected companies. Based on the interventions and evaluations from these cases, a final beta version of the artefact will be developed. The research team has initiated the activities of this cycle and we are currently applying the reference model in several companies.

Reflection and learning is conducted in parallel to the first two stages. It includes the continuous reflection on the design and redesign of the artefact as well as the documentation of requirements and the detailed descriptions of the reference model and its design areas.

Formalization of learning ensures that learnings from company-specific instantiations are further developed and documented as general solutions. In addition to general recommendations on data management in the digital economy, we are preparing a formal description of the reference model in the form of a meta-model.

4 The Data Excellence Model – Reference Model for Data Management in the Digital Economy

4.1 Purpose and Requirements

The reference model aims at structuring the main design areas of effective data management, while – at the same time – addressing the requirements of the digital and data-driven economy. These requirements have been derived from the focus groups and are supported by literature. They are summarized in **Fehler! Verweisquelle konnte nicht gefunden werden.**. Data are business-critical in the digital economy. Identifying and addressing data needs of the business requires – apart from technical capabilities – close alignment between data management and the business as the consumer of data-driven insights (R1) [31]. R2 refers to the growing number of digital services for business and private purposes – such as smart factories, smart products or social media – that increases the number of data sources and the volume of data available. For making use of big data and generating data-driven insights, data management needs to expand its traditional scope on master and transactional data to include further data types – like meta, analytical, or sensor data [32]. R3 is motivated by the high portion of compliance-, privacy-, and security-critical data created in the digital economy [33]. Consequently, not only data quality but also these further aspects have to be taken into account by data managers. Finally, as the importance and scope of data management grow in the light of the digital economy, investments are required. To justify these investments, the value generated by data and the contribution of data management to the business activities need to be transparent (R4).

Table 1. Requirements of the Digital Economy on Data Management

Requirement	Description	Design Decision
R1: Address the increasing business criticality of data	Identification of business-critical data needs	Introduce the "goals" that translate business capabilities into data management capabilities
R2: Manage data from different sources and for different purposes	Inclusion of further data sources and types (e.g. meta, analytical, or sensor data) in addition to master and transactional data	Introduce data lifecycle as "enabler" and implicitly address further data types in all other "enablers"
R3: Address relevant data-related concerns	Focus on data compliance, data privacy and security in addition to data quality	Introduce data excellence as "result" of data management that covers quality, compliance, privacy, and security
R4: Demonstrate the value contribution of data management to business	Transparency about the value contribution of data management to the business	Introduce business value as "result" of data excellence contributing to processes, customers, financials, learning and growth

Model Overview

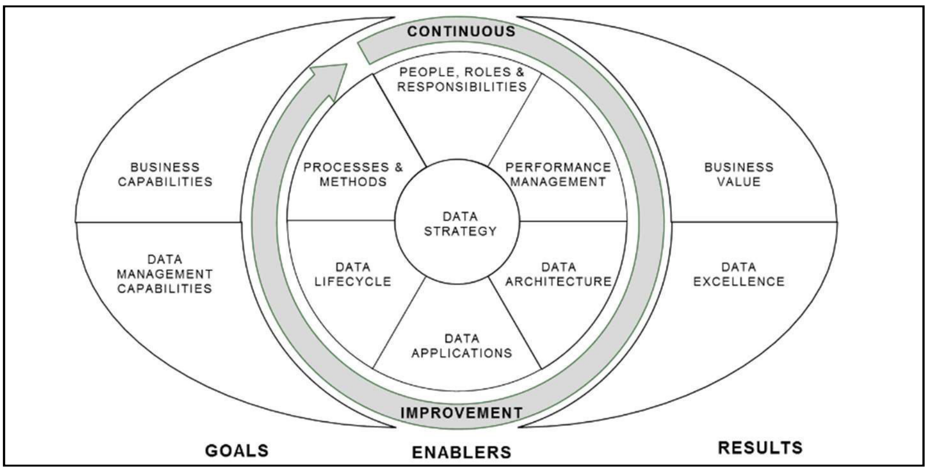

Fig. 2. Reference Model for Data Management in the Digital Economy

The first BIE cycle resulted in an alpha version of the reference model for data management in the digital and data-driven economy. **Fehler! Verweisquelle konnte nicht gefunden werden.** depicts this version. The descriptions of the design areas are detailed in **Fehler! Verweisquelle konnte nicht gefunden werden.**.

4.2 Structure of the Reference Model

Given the understanding of data as a strategic resource for the digital economy, the structure of the reference model builds on existing work from performance management, which measures, controls, and communicates indicators to improve the organizational achievement of objectives [34]. Performance management approaches consider a continuous management cycle with four phases, which are often referred to as plan, do, check, act. The reference model reflects this. It organizes design areas for data management in three categories – **goals**, **enablers**, and **results** – that are interlinked in a **continuous improvement** cycle. Goals define the strategic direction for data management. Enablers facilitate the goals. Results measure the achievement of the goals. Improvement emphasizes the dynamic nature of the model, indicating a process to adjust the goals and improve the enablers.

Design Areas

As a starting point for defining the design areas of the reference model, we reviewed the elements of existing data management models (cf. [11–22]) and data management literature. This review identified the following most common design areas of data management that we consider as "enablers": data strategy, performance management, organization, processes and methods, data architecture, and data management applications [35]. These design areas were confirmed as highly relevant by the data managers in the consortium research program. However, in order to address the requirements of the digital economy (see **Fehler! Verweisquelle konnte nicht gefunden werden.**), we consider – based on the practitioners' input – five further design areas. For ensuring the alignment of data management with the business (R1), we introduced capabilities in the goals section of the model. Capabilities describe what a company does or it should be doing [36]. By first reviewing "business capabilities" and, then, identifying the required "data management capabilities", data managers are able to directly align with the business. For the new data types introduced by digitalization (R2), we realized that they influence almost all other design areas. However, we added "data lifecycle" as an additional enabler to reflect the involved data managers' need for documenting and reviewing sources, operational activities, consumers, and purposes of data. The outcomes of data management are twofold. First, data management has a direct impact on data itself, defined as "data excellence" in the reference model. These data-related results consider data quality levels as well as the fulfillment of data compliance, data security, or data privacy requests (R3). Second, data excellence contributes to creating value to the business (R4). The “business value” design area reflects this. In line with the dimensions of the Balanced Scorecard [37], the impact of data management on the company’s financials, business processes, customers, and growth potential is reviewed.

Table 2. Descriptions of the Design Areas

Design Area	Description
Business Capabilities	are sets of skills, routines, and resources a company needs to have in order to achieve business objectives.
Data Management Capabilities	are sets of skills, routines, and resources a company needs to have in order to support business capabilities through data management.
Data Strategy	defines the scope and objectives of data management and specifies the roadmap for providing the data management capabilities required.
People, Roles and Responsibilities	defines the skills and organization to ensure effective data management and consistent use of data across the entire organization.
Performance Management	defines the measures to monitor and control the performance (i.e., progress and outcome) of data management with the help of a key performance indicator system.
Processes and Methods	defines procedures and standards for managing and using data properly and consistently.
Data Architecture	defines the conceptual data model, specifies which data is stored in which application, and describes how data flows between applications.
Data Lifecycle	defines data objects and documents, and reviews data sources, operational data activities (i.e., ranging from data acquisition and creation to data archiving), data consumers, and data use contexts.
Data Applications	defines the software components supporting data management activities.
Data Excellence	refers to the impact of data management on the data itself, first and foremost with regard to data quality (defined as “fitness for purpose”), but also with regard to additional data related aspects, such as data compliance, data security and privacy.
Business Value	refers to the impact of data management on business with regard to financials, business processes, customers, and organizational growth.

5 Evaluation

For evaluating the alpha version of the artifact, we conducted a questionnaire-based evaluation following the criteria presented by Prat et al. [38]. 25 experienced data managers participated in the survey, which comprised 24 five-point Likert-scale and seven open questions. After a presentation of the reference model, participants were asked to evaluate the structure (i.e. the completeness, simplicity, clarity, style, homomorphism, level of detail, consistency), the adaptability (i.e. robustness, learning capability), and the environmental fit (i.e. personal and organizational utility, understandability, organizational fit) of the reference model for data management in the digital economy. Overall, 86 percent of the respondents confirmed that the reference model is useful for their data management activities. On a more detailed level, 88 percent regard the reference model as complete (i.e. it covers all relevant areas), 83 percent agreed that the model depicts the reality of data management, and 80 percent regard it as robust enough to reflect future changes in the environment of data management.

First instantiations also confirm the utility of the artefact. For example, we applied the reference model for developing a data management strategy for a European-based healthcare company by first discussing the strategic objectives of the company, identifying required business capabilities, and deriving the necessary data management capabilities. We, then, reviewed the status of the six enablers, developed a target state for every design area of the enablers, and defined data excellence and business value metrics to measure and control the progress and performance of data management.

6 Conclusion and Outlook to Further Work

The paper presents a reference model for data management in the digital economy that was systematically developed in the four stages proposed by ADR. The emerging artefact provides a reference for structuring, reviewing, and establishing the design areas of data management. The evaluation of the alpha version of the artifact as well as first instantiations have demonstrated its utility. Limitations of the artefact stem from the consortium program as the research context. This program comprises only companies with a European origin. Furthermore, the data management activities of these companies have generally a high maturity and the participants share a common understanding of data management through a longtime membership in the program and frequent interactions on five workshops per year.

Further research activities are currently ongoing to instantiate the artifact in further companies and refine the reference model in the second BIE cycle. Planned results of these research activities include instantiations as well as more detailed descriptions of the scope and deliverables for every design area. To the best of our knowledge, the presented reference model is one of the first systematic approaches to extend data management in order to cover the requirements of the digital and data-driven economy.

References

1. Otto B, Auer S, Cirullies J et al. (2013) Industrial Data Space: Digital Souvereignity Over Data. Fraunhofer-Gesellschaft
2. Kagermann H, Wahlster W, Helbig J (2013) Recommendations on implementing the strategic initiative Industrie 4.0: Final report
3. Mosley M, Brackett M, Earley S et al. (2009) The DAMA Guide to the Data Management Body of Knowledge (DAMA-DMBOK guide), First edition
4. Fettke P, Loos P (eds) (2007) Reference Modeling for Business Systems Analysis. IGI Global
5. Wang RY (1998) A Product Perspective on Total Data Quality Management. Comminications of the ACM 41(2): 58–65
6. Otto B (2011) Quality Management of Corporate Data Assets. In: Praeg C-P, Spath D (eds) Quality Management for IT Services. IGI Global, pp 193–209
7. Falge C, Otto B, Österle H (2013) Towards a Strategy Design Method for Corporate Data Quality Management. In: Proceedings of the 11th International Conference on Wirtschaftsinformatik, pp 801–815

8. van den Hoven J (1999) Information Resource Management: Stewards of Data. Information Systems Management 16(1): 88–90
9. Ofner MH, Straub K, Otto B et al. (2013) Management of the master data lifecycle: A framework for analysis. Journal of Ent Info Management 26(4): 472–491
10. Otto B, Hüner KM, Österle H (2012) Toward a functional reference model for master data quality management. Inf Syst E-Bus Manage 10(3): 395–425
11. Mosley M (2008) DAMA-DMBOK Functional Framework: Version 3.02
12. Otto B, Österle H (2016) Corporate Data Quality: Prerequisite for Successful Business Models. Springer Berlin Heidelberg, Berlin, Heidelberg
13. Thomas G (2014) The DGI Data Governance Framework
14. Capgemini (2010) Data Governance & Stewardship
15. Radcliffe J (2007) The Seven Building Blocks of MDM: A Framework for Success
16. Yuhanna N, Leganza G, Karel R et al. (2011) Forrester's Data Management Reference Architecture: Align Data Management Roles To Create Stronger Business Value
17. Sivaprkasam SR (2010) Enterprise Data Management Approach: A Comprehensive Data Approach for CSPs
18. Sun H, Xu S (2012) Oracle Enterprise Architecture Framework: Information Architecture Domain
19. IBM Corporation (2007) The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance
20. Informatica (2013) Holistic Data Governance: A Framework for Competitive Advantage
21. SAS (2014) The SAS Data Governance Framework: A Blueprint for Success
22. GS1 (2010) GS1 Data Quality Framework: Version 3.0
23. Batini C, Cappiello C, Francalanci C et al. (2009) Methodologies for data quality assessment and improvement. ACM Comput. Surv. 41(3): 1–52
24. Matt C, Hess T, Benlian A (2015) Digital Transformation Strategies. Bus Inf Syst Eng 57(5): 339–343
25. Chen J, Chen Y, Du X et al. (2013) Big data challenge: A data management perspective. Front. Comput. Sci. 7(2): 157–164
26. Atul S, Desale Girish B, Patil Swati P (2017) Big Data and Big Data Management (BDM) with current Technologies –Review. IJERA 07(04): 42–44
27. Chen H, Chiang RHL, Storey VC (2012) Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly 36(4): 1165–1188
28. Gregor S (2006) The Nature of Theory in Information Systems. MIS Quarterly 30(3): 611–642
29. Österle H, Otto B (2010) Consortium Research: A Method for Researcher-Practitioner Collaboration in Design-Oriented IS Research. Bus Inf Syst Eng 2(5): 283–293
30. Sein MK, Henfridsson O, Purao S et al. (2011) Action Design Research. MIS Quarterly 35(1): 37–56
31. Proper H, Lankhorst MM (2014) Enterprise Architecture: Towards essential sensemaking. Enterprise Modelling and Information Systems Architectures 9(1): 5–21
32. Lee YW, Madnick, Stuart E. et al. (2014) A Cubic Framework for the Chief Data Officer: Succeeding in a World of Big Data. MIS Quarterly Executive 13(1): 1–13
33. Sadeghi A-R, Wachsmann C, Waidner M (2015) Security and privacy challenges in industrial internet of things. In: Proceedings of the 52nd Design Automation Conference, pp 1–6
34. Fischer TM, Möller K, Schultze W (2012) Controlling. Schäffer-Poeschel, Stuttgart
35. Labadie C, Legner C Overview of Data Management Frameworks
36. Bērziša S, Bravos G, Gonzalez TC et al. (2015) Capability Driven Development: An Approach to Designing Digital Enterprises. Bus Inf Syst Eng 57(1): 15–25

37. Kaplan RS, Norton DP (1996) The balanced scorecard: Translating strategy into action. Harvard Business Review Press, Boston
38. Prat N, Comyn-Wattiau I, Akoka J (2015) A Taxonomy of Evaluation Methods for Information Systems Artifacts. Journal of Management Information Systems 32(3): 229–267