International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:11, No:4, 2017

# Conceptualizing the Knowledge to Manage and Utilize Data Assets in the Context of Digitization: Case Studies of Multinational Industrial Enterprises

Martin Böhmer, Agatha Dabrowski, Boris Otto

***Abstract*—**The trend of digitization significantly changes the role of data for enterprises. Data turn from an enabler to an intangible organizational asset that requires management and qualifies as a tradeable good. The idea of a networked economy has gained momentum in the data domain as collaborative approaches for data management emerge. Traditional organizational knowledge consequently needs to be extended by comprehensive knowledge about data. The knowledge about data is vital for organizations to ensure that data quality requirements are met and data can be effectively utilized and sovereignly governed. As this specific knowledge has been paid little attention to so far by academics, the aim of the research presented in this paper is to conceptualize it by proposing a "data knowledge model". Relevant model entities have been identified based on a design science research (DSR) approach that iteratively integrates insights of various industry case studies and literature research.

***Keywords*—**Data management, digitization, Industry 4.0, knowledge engineering, metamodel.

## I. Introduction

In recent years, enterprises are facing a set of new business drivers that are based on technological advance which can be summarized by "digitization". Customer demand is emphasizing individualization, on-demand (service) delivery, and the combination of physical products with digital services, e.g. smart watches [1]-[3]. Business is increasingly driven by (big) data and facing the automation of decision making [4], [5]. A significant share of decision making is shifted "down" to the device level by the emerging Internet of Things (IoT) and Industry 4.0 [2], [6]. The quality of decisions mainly depends on the ability to consolidate insights from different data sources and the ability to detect and integrate different levels of data quality [5], [7]. These challenges consequently demand enterprises to develop their data management capabilities [8] in order to (1) identify and actively manage data assets, (2) clearly define and exercise data ownership, (3) provide means for semantic integration and (4) treat data as a product involving a managed data supply chain. Hence, the perception of data has changed from an enabler to an asset that has a distinct business value and requires dedicated management [9]. In contrast to physical assets, the management of data assets needs to be reconsidered as new technologies for data management emerge. By enabling fast processing of large datasets in-memory databases have proven to create tangible business benefits [4]. Data spaces address the management and integration of plenty of heterogeneous data sources [10].

Company boundaries increasingly blur in the area of data management as data turns into a good that is shared and collaboratively maintained among enterprises [11], [12]. An important implication of a networked "data economy" [9] is that it is not only the actual data that need to be shared, but also the knowledge required to source, integrate, utilize, maintain, govern, and control it. Such "data knowledge" provides answers to fundamental questions, e.g. about its meaning, structure, or creator/provider. A key requirement for collaborative data management approaches is that they maintain the sovereignty of the involved parties over their data [11]. This requires knowledge to be shared about the governance of the data, like its integrity, rules, ownership, and validity.

Although data knowledge is highly relevant for the digitization of the industrial enterprise, there is little insight about its constituents. Researchers and practitioners require a common and detailed understanding of what data knowledge is in order to master the challenges outlined above and build innovative concepts on top of it. Hence, findings from related fields of research, like data management, enterprise architecture and metadata management, need to be consistently integrated and complemented by new insights. This motivates the guiding question for the research presented in this paper: What constitutes data knowledge? The central research objective is to design a conceptual model that summarizes relevant data knowledge (referred to as "data knowledge model"). This objective is approached by three research questions (Sub-RQ):

1) Sub-RQ1: What is a general definition of data knowledge?
2) Sub-RQ2: What is the specific role of data knowledge in an Industry 4.0 context?
3) Sub-RQ3: What are relevant entities to describe the structure and semantics of data knowledge?

Section II provides the theoretical background for our research, followed by a description of our research method outlined in Section III. A definition of data knowledge is

M. Böhmer is with the Fraunhofer Institute for Material Flow and Logistics, Dortmund, Germany (phone: +49 231/9743-203; fax: +49 231/ 9743-77-203; e-mail: martin.boehmer@iml.fraunhofer.de).

A. Dabrowski is with the Fraunhofer Institute for Material Flow and Logistics, Dortmund, Germany (e-mail: agatha.dabrowski@ iml.fraunhofer.de).

B. Otto is with the Audi-Endowed Chair of Supply Net Order Management, TU Dortmund, Germany (e-mail: boris.otto@tu-dortmund.de).

proposed in Section IV (Sub-RQ1). Particular attention is given to the Industry 4.0 context in Section V (Sub-RQ2). Our findings are presented in Section VI by sharing insights from the design phases and presenting the resulting data knowledge model (Sub-RQ3). We evaluate our findings in Section VII. Section VIII concludes the paper.

## II. Background and Related Work

This chapter introduces the fundamental concepts. Works directly related to the data knowledge model are reviewed in the course of the design process (Section VI.*B*).

### *A. Data Management*

The leading practitioner organization DAMA, the Data Management Association, defines data management as "the business function of planning for, controlling and delivering data and information assets" [13]. It comprises several functions of which we name those relevant for our research. Data governance is the cross-cutting function of planning and controlling data, its use and its management. Data quality management defines appropriate metrics to assess and monitor data quality. Data architecture management identifies and transforms (business) requirements into data models and different types of architectures that meet those requirements. Metadata management supports other data management functions by integrating and providing information (metadata) required to perform them.

The Framework for Corporate Data Quality Management (CDQM) is an approach to data management in multinational companies that comprises six design areas [14], [15]. The areas of strategy, controlling, organization, processes, data architecture as well as applications are intended to holistically cover data management.

In addition to these two major frameworks, various other frameworks have been developed for data (quality) management and data governance [16]-[19]. They predominantly originate from business practice or consulting and mainly differ in terms of wording or in emphasizing a particular detail or perspective.

### *B. Knowledge Management*

Strategic management of knowledge is considered to be the most significant source of organizational competitive advantage in an increasingly dynamic and rapidly changing environment [20], [21]. Knowledge management in an organizational context originated from strategic management [22] and builds a knowledge-based perspective on top of the resource-based view of a company [23], [24]. The knowledge-based view (KBV) argues that the (successful) combination and application of organizational resources mainly relies on a firm's knowledge and the ability to apply it [24]. It is influenced by several other areas, such as information economics, the organizational environment (e.g., culture, structure, behavior) and quality management [25].

The basic processes of knowledge management comprise the creation and maintenance, storage and retrieval, transfer and sharing, as well as the application of knowledge [23]. Our data knowledge model is intended to describe the subject matter of these processes in the data management context and to enable its processing in knowledge management systems (KMS) [26].

## III. Research Method

### *A. Methods and Techniques*

DSR has been chosen as the foundational methodology for this research because it aims at developing new and useful artifacts for research and practice [27]. A common type of DSR artifact suitable for approaching our research questions is a model comprising the basic constructs of a domain and the relationships among them [28]. Consortium research has proven to be useful for ensuring relevance in DSR by facilitating the close interaction between academic researchers and a number of partner companies [29].

Our research environment is a consortium with 20, mainly multinational partner companies. A wide range of industries is represented in the consortium showing a slight bias towards manufacturing and process industries. Data managers and architects from this consortium identified the lack of a specific knowledge conceptualization as a severe impediment. In terms of the DSR terminology, our research emerged from an objective-centered initiation [27] aiming at the development of an artifact that contributes to this request, our research question(s) and information systems research.

The fact that the partner companies were highly interested in contributing to the research provided the opportunity for case study research (CSR). The aim of CSR is to analyze complex subject matters in a defined, practical environment without the necessity of prior scientific results, such as literature or empirical evidence [30]. Participatory CSR merges the subject matters of the research and the case study by enabling researchers to influence actions taken and observe their effects [31]. In the course of our design-oriented approach, all types and use cases of CSR were employed.

Our research is further based on a range of scientifically established design and research techniques; namely, reference modeling [32], structured interviews [33], and literature research [34]. In the course of the case studies, interviews were conducted with individuals and groups of employees from the respective company with different backgrounds and responsibilities. Additional interviews were run for a broader evaluation at multilateral practice workshops with participants from several partner companies of the research consortium (so-called focus group interviews).

### *B. Research Process*

Acknowledging the principles and recommendations from the chosen research methods and techniques, our research process has five phases comprising of a series of case studies. Each phase iteratively advances the targeted data knowledge model in that it builds on the findings of the previous phases. The results of the first two phases were evaluated by employing focus group interviews to ensure their relevance. Putting emphasis on findings from CSR, we chose an inductive approach to our research questions to be able to closely exchange with our research consortium. Theoretical considerations conclude our process to reflect our findings based on related works from scientific literature.

1. Case Study “Logistics Mall”

The research project “Logistics Mall – Cloud Computing for Logistics” aimed at the development of a cloud-based platform for modeling and executing logistics processes based on modular, self-contained application building blocks (called “apps”) from an open marketplace [35]. It qualified for contributing to our research due to the key role of data as means for interoperability and the collaborative setting to manage, share and utilize data in a cloud environment [36].

Six semi-structured interview have been conducted with different stakeholders of the platform between Aug 2012 and Nov 2013. These involved a total of 20 participants representing the platform operator, app providers and users of the platform, like process modelers and app users. Emphasis has been put on Sub-RQ3 by asking for relevant metadata concerning the master data and transactional data processed by the platform.

The initial version of the model that resulted from this case study was evaluated by a focus group interview in Dec 2014 with 11 participants of partner companies from the pharma, consulting, automotive, engineering, transportation, telecommunications and energy industry.

2. Case Study at Major Global Pharma Enterprise

This case study examined a project of the Marketing and Sales function at a global pharmaceutical enterprise that was driven by the objective to better integrate business processes across over 100 country divisions. The project was selected due to the fact that operational process issues in the considered business function mainly resulted from a lack of knowledge about data. Examples are unclear data dependencies and data qualities requirements of the processes in scope.

Four semi-structured interviews have been conducted between Mar and Apr 2015 involving a total of 24 participants. The interviewees had diverse backgrounds from marketing, sales and IT and spanned different hierarchy levels of the business function. Sub-RQ1 and Sub-RQ3 were covered by this case study to advance our data knowledge model and derive a first definition of data knowledge.

Phase two concludes with another focus group interview attended by nine representatives of partner companies from the pharma, software, telecommunications, and automotive industry.

3. Participatory Case Study at Major Global Engineering Corporation

In this case study, we accompanied a strategic initiative at a major global engineering company to establish a knowledge base for corporate-wide data quality management. The objectives were twofold and contribute directly to our research questions. Firstly, the distributed knowledge about data should be integrated and made centrally available. Secondly, the transfer of data knowledge between employees should be facilitated.

Between Jun and Dec 2015, 22 interviews were conducted with a total of 30 participants representing different business divisions (subsidiaries), functions, hierarchy levels, and data management scopes (global, divisional, functional and regional). Due to the available time and the small groups interviewed, all research questions could be considered.

4. Participatory Case Study at Global Automotive Supplier

This case study was conducted in the course of a strategic program to establish data governance and data quality management as a corporate function. It emphasizes the particular complexity of a multi-national industrial enterprise as global, functional, divisional and regional requirements need to be aligned and specific knowledge about data is required at each respective level.

A total of six people were individually interviewed based on a semi-structured approach. Although all research questions were covered, emphasis has been put on Sub-RQ3 in order to advance our model.

5. Analysis of Scientific and Practical Literature

The literature review concludes our research. On the one hand, it aims to gather the scientific state of the art to incorporate it into the data knowledge model. On the other hand, it serves to evaluate the practical findings from the case studies. As a result, all research questions have been considered during this phase. However, special attention has been given to Sub-RQ1 to link the concept of data knowledge to common concepts from literature.

The taxonomy for literature reviews by Cooper [37] was used to specify our literature research: The focus is on research outcomes with the intention of integrating them. Considering our objective-centered DSR approach and the practically-oriented research environment, we focus on the identification of pivotal works.

The Scopus database was chosen as it is the largest database of peer-reviewed literature. Results sets were iteratively narrowed down based on relevance of titles, abstracts and full texts according to vom Brocke et al. [34]. In addition, the experience embodied in our research consortium was leveraged by asking fellow researches and practitioners for related works. A forward and backward search was conducted on the suggested contributions to identify pivotal ones.

## IV. Data Knowledge

Data, information, and knowledge can be related hierarchically in terms of complexity [13], [23], [38], [39]. Data represent the least complex level serving as abstractions of real facts that can be stored and processed [39]. Similar to any other corporate resource, it has its own lifecycle comprising seven phases: plan, specify, enable, create & acquire, maintain & use, archive & retrieve, and purge [13]. Due to the fact that raw data (e.g. from sensors) are inexpressive, it needs to be enriched by a context that indicates its meaning and origin [39]. This interpretation leads to information and represents the foundation for the use of data

[13].

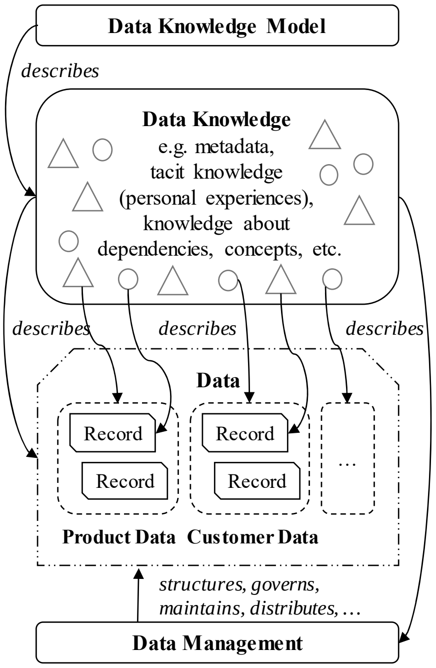


Fig. 1 The concept of data knowledge

In a corporate setting, knowledge is characterized as a tool for reasoning and decision making [23]. It is the ability to make knowledge available and apply it that enables enterprises to take better decisions and gain competitive advantage [22]. In this context, the distinction between explicit and tacit knowledge is of particular interest. Explicit knowledge represents generalized, codified knowledge, whereas tacit knowledge is incorporated in actions or experience and is therefore highly dependent on context [22].

We define data knowledge as the extensive understanding that allows an interdisciplinary, comprehensive consideration of data and information as an organizational asset. It aims to synthesize the explicit and tacit knowledge about the structural-semantic essence of data and the interrelation of its diverse contexts. Data knowledge describes and supports a distinct management of these assets, enables the value-adding utilization of data and information in business models and processes and builds the foundation to share and collaboratively maintain data assets under sovereign governance. Building on the KBV of a company, data knowledge management is referred to as an organization's capacity to perform these activities. Fig. 1 illustrates the concept of data knowledge.

Plenty of individuals and systems process data at different stages throughout the data lifecycle. Dependencies and responsibilities often remain incomprehensible from an end-to-end view, which affects consistency and business operations [14], [15]. Addressing the complexity of this setting requires an interdisciplinary perception of data knowledge that considers the requirements of different corporate functions, units and stakeholders along the entire data lifecycle. Due to the fact that data knowledge is as distributed as the actual data, the conceptualization of data knowledge needs to provide means to consistently integrate sources of knowledge that each contains fragments of an overall picture of a data asset. This goes beyond the mere integration of metadata.

TABLE I
RAMI4.0 SECTIONS CONSIDERED FOR OUR DATA KNOWLEDGE MODEL (HIGHLIGHTED CELLS)

| Dimension | Architecture Layer | Life cycle & value stream | Hierarchy |
|---|---|---|---|
| Sections (to be read column-wise) | Asset | Type | Product |
| | Integration | | Field device |
| | Communication | | Control device |
| | Information | | Station |
| | Functions | Instance | Work Center |
| | | | Enterprise |
| | Business | | Connected world |

## V. DATA KNOWLEDGE IN THE INDUSTRY 4.0 CONTEXT

Automation, decentralization, autonomy (also referred to as smartness or self-organization), and interconnection are the key principles of the networked, cyber-physical systems that blur the lines between human and non-human actors and that are referred to as Industry 4.0 (I4.0) [2], [40]. These systems essentially represent a composition of distinct, self-contained assets. DIN SPEC 91345 defines such an asset as an intentionally created artifact that is designated to take a particular role in a system and thus has a value to an organization across its entire lifecycle [41]. As the combination of assets results in a new asset with distinct features, a distance measuring sensor, a production machine, and an entire plant are all considered individual assets. In consequence, a key challenge for I4.0 is to achieve universal interoperability among all these assets. Standards and reference models are essential means in this regard.

The Reference Architecture Model for Industry 4.0 (RAMI4.0) aims to provide a three-dimensional description of an asset [41], [42], as outlined by Table I. Interoperability is approached by introducing a so called Administration Shell that provides a standardized, digital description of an asset according to the different RAMI4.0 dimensions [41], [43]. The Administration Shell is a virtual representation and encapsulation of an asset that enables it to interact with other I4.0 components in unified ways. Its inner structure comprises a header and a body part.

While the body comprises detailed information about the asset (e.g. technical specification, functionality) and data created by the asset (e.g. sensor data), the header functions as a "table of contents" (manifest) and an active, external interface (the component manager) for the body and the asset. The body is further organized into diverse partial (information) models made of hierarchically structured, coherent properties that incorporate a particular purpose and relate to various RAMI4.0 sections. Consequently, the

Administration Shell is able to describe any facets of an asset and functions as a framework for future (content) reference models.

Revisiting the I4.0 concepts outlined above from a data management perspective illustrates the essential role of data knowledge for I4.0 scenarios and reveals the need for further research. As delineated in the introduction, data has turned into an organizational asset itself. The I4.0 asset definition of DIN SPEC 91345 consistently includes data as an asset. Similar to physical assets, data assets are intentionally created or captured to serve a particular purpose in some system. The fact that data have become a tradable good [9], [44] and are subject to systematic quality management [14], [15] proves its organizational value. As a result, data have an equal importance for I4.0 scenarios as physical assets. In contrast to physical assets, there is little insight into the management of data assets in these scenarios, which motivates research for means of compatible integration.

Given that data are inherently passive, it needs to be complemented by an Administration Shell to become an I4.0 component. It is what we consider data knowledge that constitutes the content of such an Administration Shell to describe the data asset from various perspectives. In this regard, the data knowledge model presented in this paper represents a partial model that covers particular sections of the RAMI4.0 (Table I and Section VII for a discussion).

Although I4.0 shifts the ways that organizations operate, from a data standpoint, there is continual "nucleus data" that all operations inherently rely on and that essentially distinguishes an organization from another (e.g., data of customers, products, orders, inventory) [5]. Hence, it is the integration of nucleus data with new types and sources of data as well as the adoption of adequate data management principles that shape the I4.0 transformation on the information layer of the RAMI4.0. This process consequently requires sound data knowledge about the essential, continual aspects of nucleus data management and motivates the scope of our research. Our data knowledge model aims at conceptualizing this knowledge.

## VI. The Data Knowledge Model

### A. Design Results

The data knowledge model is a conceptual model covering relevant entities for data knowledge to describe the structure and semantics of data. It is structured into seven areas. *Business Data Definition* is the core area and is complemented by *Governance*, *Change*, *Operations*, *Data Quality*, *Applications,* and *Support* that each outlines a specific context (Fig. 2). Each area involves a set of entities (subsequently marked in bold) that represent either explicit or tacit knowledge. Table II specifies the relationships among the entities and provides a detailed view of the model.

The area *Business Data Definition* comprises the structural and semantic data knowledge that describes the subject matter (data). A **business object type** (BOT) outlines a conceptual (data) entity that represents a coherent set of information about a group of real or virtual objects (e.g., customer, bill of material, sales region). It its characterized by its **business object type attributes** (BOTAs), conceptually elementary pieces of information (e.g. tax ID, address). BOT and BOTA can both be illustrated by an **example** to indicate their valid and invalid use, whereas only BOTAs can refer to a **value list**, i.e. lists with pre-defined values like country codes and incoterms. A **business object** (BO) represents a specific, individual instance of a BOT (e.g. a certain supplier "Fluid Supply Ltd"). Similarly, a **business object attribute** (BOA) is a particular instance of a BOTA, e.g. the tax ID "DE-123456-789". **Business object domains** (BODs) describe groups of BOTs either from a subject matter view or from a management view. A subject matter view considers the functional coherence of BOTs, e.g. product, business partner or marketing. In contrast to that, a BOD from a management view regards BOTs associated with certain management principles or (quality) requirements, e.g. customer-facing vs. internal BOTs or regional BOTs.

Knowledge required to govern, control, and manage data is covered by the area *Governance*. An **organizational unit**, for instance a department, represents a constituent of the company's organizational structure. It provides hierarchical context for a specific **person,** i.e. an individual assigned with certain roles. **Role** entities represent abstract definitions of a particular (data-management-related) function including its goals and authorities. Furthermore, roles can take **responsibility** for certain scopes and are influenced by internal or external regulatory drivers, conceptualized as **regulation** (e.g. directives, policies, laws, etc.). The lifecycle stages of a BOT and hence the states of their instances (BOs) are defined as the **data lifecycle** and emerge from the *Governance* area as well.

Aspects of change to either data or data management are subsumed and structured by the *Change* area. The central concept of this section is a **project** implying these aspects. To further characterize it, a **project plan** illustrates when a particular action is scheduled. In this context, **project experience** is of particular value. This type of tacit knowledge supports problem solving and decision making for strategic and conceptual data management issues. It includes the knowledge of success factors, involved risks as well as the ability to perform a sound evaluation of a solution or decision. Examples are the implementation of a regulatory requirement, the causal identification and resolving of data defects or the integration of data after merges and acquisitions. Another relevant type of tacit data knowledge is **social competence**. Interaction among stakeholders is an important success factor for data management in large enterprises. In this regard, tacit knowledge includes insights on how to approach individuals or groups to convince or motivate them. Especially when intending to change data or data management, social competence plays a considerable role. A **change request** is a detailed description of a specific change requirement.

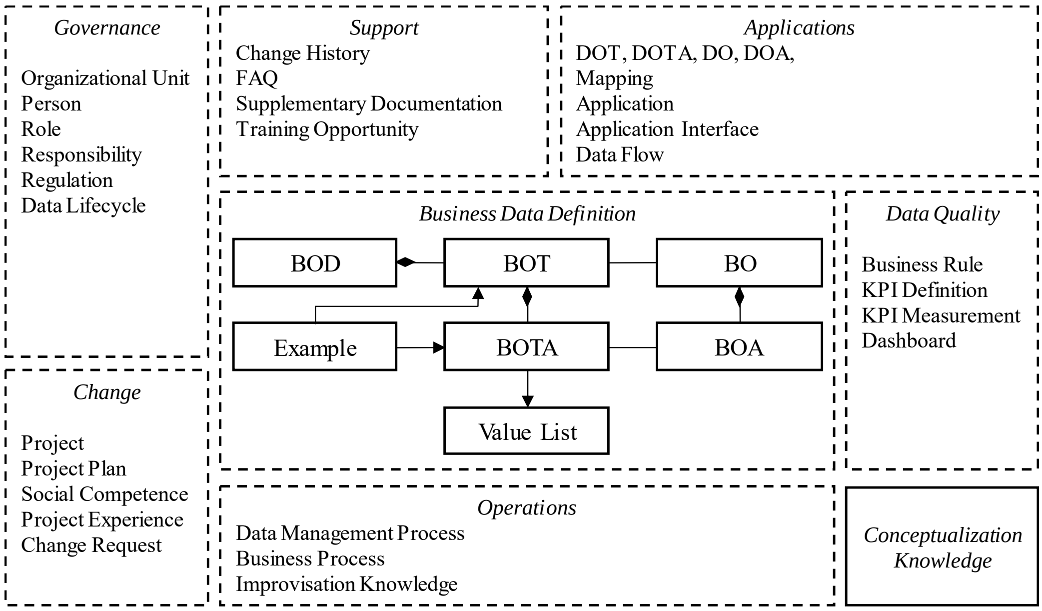


Fig. 2 High-level overview of areas and selected entities

The area of *Operations* contains process-related data knowledge for business operations as well as data management operations. A **business process** describes a sequence of activities consuming or relying on data and represents a data usage context. In contrast to that, a **data management process** describes a sequence of activities for data maintenance or governance. **Improvisational knowledge** is important for both of these processes, as it resembles tacit knowledge to handle operational challenges or exceptions. In contrast to project experience, improvisational knowledge mainly enables the users of data to handle errors and exceptions. This type of knowledge also facilitates the "invention" of workarounds. Due to the complex interrelations of cause and effects between processes, data and applications, this type of knowledge is crucial for maintaining business critical operations. Examples for such knowledge are the ability to correct data defects or to manually perform automated tasks.

Quality aspects concerning data are collected in the area of *Data Quality*. A specific **business rule** defines the integrity of a single BO or across them, e.g. the mandatory provision of a tax ID. Based on a set of these rules, a **KPI definition** characterizes abstract data quality measures, whereas a **KPI measurement** represents the actual performance values measured at a particular time. Sets of these measurements can be gathered for a particular purpose and collected in a **dashboard** or report.

The area of *Applications* conceptualizes the technical representation, the storage and the flow of data and thus represents data lineage and data distribution knowledge. Data are processed and stored in an installed instance of an **application** by means of specific, technical counterparts of the concepts outlined in the *Business Data Definition*. Since the business object type is a conceptual model, it needs to be implemented according to the application system it is stored in. A **data object type** (DOT) resembles an application-specific, technical representation of a BOT. In this context, a **data object type attribute** (DOTA) is an individual atomic field to store a piece of information, e.g. a certain database column. Similarly, the individual record of a DOT is represented as a **data object** (DO), whereas a **data object attribute** (DOA) is a piece of information being a part of a DO, e.g. the certain value of a record's database column. These implementations vary in complexity, focus and quality depending on the application they are stored in or processed with. BOTs and DOTs represent different levels of abstraction in data architecture that are commonly referred to as conceptual data or business object models (BOT) and logical or technical data models (DOT) [13], [45]. To be able to retrieve the most suitable datasets for a certain business object type, consistent **mapping** from abstract model (BOT(A)) to implemented model (DOT(A)) is important. An **application interface** is a technical specification of an application's means to exchange data. The actual connection and data exchange between particular applications is indicated by **data flows**.

The area *Support* subsumes supplementary aspects that complement the other areas and facilitate the dissemination of all types of data knowledge. To assist this objective, a **change history** may contain information that allows to track the development of data definitions, processes and many more (e.g., field "age" added to customer). A collection of frequently asked questions (**FAQ**) manages to complement the supportive effect by expressing explicit or tacit elements of data knowledge. In addition, material of **supplementary documentation** such as presentations or videos can function as a reference or learning capability for data management operations. Furthermore, it is possible to schedule a **training opportunity** as a planned event a person can attend.

TABLE II
ENTITIES OF THE DATA KNOWLEDGE MODEL AND THEIR RELATIONSHIPS

| | Application | Application Interface | Business Object | Business Object Attribute | Business Object Domain | Business Object Type | Business Object Type Attribute | Business Process | Business Rule | Change History | Change Request | Conceptualization Knowledge | Dashboard | Data Flow | Data Lifecycle | Data Management Process | Data Object | Data Object Attribute | Data Object Type | Data Object Type Attribute | Example | FAQ | Improvisation Knowledge | KPI Definition | KPI measurement | Mapping | Organizational Unit | Person | Project | Project Experience | Project Plan | Regulation | Responsibility | Role | Social Competence | Supplementary Documentation | Training Opportunity | Value List |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Application | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Application Interface | X | | | | | | | | | | | | | X | | | | | | | | | | | | | | | | | | | | | | | | |
| Business Object | | | X | X | | X | X | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Business Object Attribute | | | X | X | | | X | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Business Object Domain | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Business Object Type | | | X | X | X | X | X | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Business Object Type Attribute | | | | | | X | X | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| Business Process | | | X | X | | X | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Business Rule | | | X | | | | | | | | | | | | | | | | | | | | | X | | | | | | | | | | | | | | |
| Change History | X | X | X | X | X | X | X | X | X | | X | | X | X | X | X | X | X | X | X | X | X | | X | X | X | X | X | X | | X | X | X | X | | X | X | X |
| Change Request | | | | | | | | | | | | | | | | X | | | | | | | | | | | | | | | | | | | | | | |
| Conceptualization Knowledge | X | X | X | X | X | X | X | X | X | X | X | | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| Dashboard | | | | | | | | | | | | | | | | | | | | | | | | | X | | | | | | | | | | | | | |
| Data Flow | | X | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Data Lifecycle | | | X | X | | X | X | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Data Management Process | | | X | X | X | X | X | | | | X | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Data Object | | | | | | | | | | | | | | | | | X | X | | | | | | | | | | | | | | | | | | | | |
| Data Object Attribute | | | | | | | | | | | | | | | | | X | X | | | | | | | | | | | | | | | | | | | | |
| Data Object Type | X | | | | | | | | | | | | | | | | X | X | X | | | | | | | | | | | | | | | | | | | |
| Data Object Type Attribute | | | | | | | | | | | | | | | | | X | X | X | X | | | | | | | | | | | | | | | | | | |
| Example | | | | | | X | X | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| FAQ | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| Improvisation Knowledge | | | | | | | | X | | | | | | | | X | | | | | | | | | | | | | | | | | | | | | | |
| KPI Definition | | | | | | | | | | | | | | | | | | | | | | | | | X | | | | | | | | | | | | | |
| KPI Measurement | | | X | X | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Mapping | | | | | | X | | | | | | | | | | | | | X | | | | | | | | | | | | | | | | | | | |
| Organizational Unit | | | | | | | | | | | | | | | | | | | | | | | | | | | X | | | | | | | | | | | |
| Person | | | | | | | | | | | | | | | | | | | | | | | | | | | X | | | | | | | X | | | X | |
| Project | | | | | | | | | | | X | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Project Experience | | | | | | | | | | | | | | | | | | | | | | | | | | | | | X | | | | | | | | | |
| Project Plan | | | | | | | | | | | | | | | | | | | | | | | | | | | | | X | | | | | | | | | |
| Regulation | X | X | X | X | X | X | X | X | X | X | X | | X | X | X | X | X | X | X | X | X | X | | X | X | X | X | | X | | X | | X | X | | X | X | X |
| Responsibility | X | X | X | X | X | X | X | X | X | X | X | | X | X | X | X | X | X | X | X | X | X | | X | X | X | X | | X | | X | X | | X | | X | X | X |
| Role | | | | | | | | X | | | | | | | | X | | | | | | | | | | | X | | X | | | | X | X | | | | |
| Social Competence | | | | | | | | | | | | | | | | | | | | | | | | | | | X | X | X | | | | | X | | | | |
| Supplementary Documentation | X | X | X | X | X | X | X | X | X | X | X | | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | | X | X |
| Training Opportunity | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Value List | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Conceptualization knowledge (CK) is not classified as part of any of the aforementioned areas. In contrast to other widely applicable entities, e.g. the change history, this type of knowledge is relevant to tacit data knowledge as well as to explicit data knowledge. Conceptualization knowledge focuses on the formulation of concepts and circumstances (e.g., of cultural nature) to support problem solving or decision making. An example for such knowledge is the ability to comprehensively describe the managed data with respect to the targeted audience or users.

### B. Contributions of Different Design Phases

- **Phase 1:** The initial case study contributes basic patterns for describing the structure and semantics of data and builds the foundation of the *Business Data Definition* area. The cloud-based platform developed by the case study project operates on a canonical data model differentiating a business view and the technical view on data definitions [36]. These views have been adopted by the separation of BOT and DOT.
- **Phase 2:** The "Governance Meta Model" (GMM) that resulted from the case study project proved to be very effective in resolving the initial business issues in the Marketing and Sales function. It is capable of describing a common, global conceptual information model (CIM) from a business perspective and link it with two contexts. Firstly, a processual context has been incorporated into the GMM by linking business processes with the attributes of the CIM that they rely on. This relationship also featured relevant governance aspects like process ownership. The second context targeted the IT applications. In order to address the operational issues and heterogeneity of systems, the mapping had to be as explicit as possible and finally reflected system instances on a release level, database tables, columns and system interfaces. While the project's CIM maps to and confirms the *Business Data Definition* area, the *Operations* and *Applications* area have added to our model to represent the processual and IT contexts of data assets.
- **Phase 3:** Since the initiative's objectives were not only integrating and making distributed explicit knowledge centrally available but also the facilitation of tacit knowledge transfer, specific types of tacit knowledge could be identified. These types of tacit knowledge are considered in our model namely as *project experience*, *improvisation knowledge*, *conceptualization knowledge* as well as *social competence*. Moreover, a clear understanding of the data management roles, their scope and implied responsibilities along with insight on the persons appointed for these roles turns out to be highly sought-after knowledge and relevant for both objectives and is therefore incorporated as well. As a result, the *Governance* area was initially added to our model.
- **Phase 4:** The Governance areas were revised by the findings from establishing data governance as a corporate function at a global automotive supplier. The main contribution of this case study originates from the activities of setting up a corporate data quality management including quality metrics and means to assess data quality. These findings have shaped the *Data Quality* of our model.
- **Phase 5:** The literature review concludes the design of the data knowledge model. A detailed comparison confirms that the data knowledge model includes key entities of related enterprise metamodels, like the Business Engineering (BE) metamodel [46], metamodels of Enterprise Architecture (EA) frameworks [47]-[49] or process-oriented frameworks for enterprise modeling [50]. These metamodels confirm our findings in that they provide means to describe data, business processes, information systems and their interrelationships.

Building on the BE metamodel, Schmidt designed a method for data integration that relies on a business data dictionary [51]. Schmidt's work contributes a conceptual differentiation of types and instances to our knowledge model (BOT/DOT vs. BO/DO) and confirms our findings from phase 2.

Automated processing of large amounts of data across distributed systems motivates the discipline of metadata management. A considerable number of standards and guidelines define sets of metadata attributes relating to data management and other contexts. We draw on the comprehensive comparison of 19 of these standards by Päivärinta et al. [52]. These insights on the attribute-level resulted in extensions of our data knowledge model, e.g. considering a change history.

## VII. Evaluation

Feedback from the conducted interviews (Section III) shows that the aspects conceptualized by the data knowledge model are practically relevant. However, their individual criticality varies generally and in specific use cases. The semantics of the model and its entities are considered to be consistent and can be linked to practical and scientific literature (Section VI.*B*).

A repeated objection concerned the representation of explicit and tacit knowledge. There is a significant difference in the specificity for data management, granularity and extent to that they have been considered. This feedback matches our impression from the literature research. Although the general relevance of tacit knowledge for data management is commonly accepted among our research partners, the value of its representation in the model turned out not to be obvious and involves additional explanation. Hence, further research is required to better understand the structure and role of tacit data knowledge, as well as to better link it with explicit knowledge in order to facilitate the value of its conceptualization.

There are some aspects of the outlined design process that result in shortcomings of the current data knowledge model and demand for further research. The case studies tend to capture an internal view on data management only. Although the designed model is expected to cover a considerable share of the cross-organizational data knowledge (management) requirements already, it lacks of explicit insights for this use case that would prove its validity. The additional consideration of appropriate use cases would eliminate this shortcoming.

Integrating our data knowledge model into the context of Industry 4.0, a classification according to the RAMI4.0 (highlighted cells of Table I) illustrates its focus on the enterprise and business unit (work center) level of data management. From a lifecycle perspective, there is a clear bias on the type level, although instances are reflected in our model. The business architecture and functional architecture layers are considered by the links that our data model maintains to business processes, data governance, and IT systems.

The conclusion from the RAMI4.0 classification is twofold. Firstly, it proves that our data knowledge model provides a significant contribution for integrating data assets in I4.0 ecosystems as several sections of the RAMI4.0 are covered. Our model particularly represents the knowledge about an organization's "nucleus data" [5] of traditional business objects that essentially distinguish the organization from another and continually drive its operations. There is an inevitable need for this knowledge in order to link nucleus data with the new sources and types of data emerging from digitization and I4.0. Secondly, knowledge about these new sources and types of data needs to be explored. Additional partial models for the Administration Shell are required to describe such (new) data assets and their relation to the present assets. Trading and sharing data also requires adequate descriptions of self-contained collections of data on the instance level to better support the idea of data as a product. The Industrial Data Space [11] and the Corporate Data League [12] are exemplary use cases for this type of data knowledge.

## VIII. Conclusion

Current developments require organizations to complement their traditional knowledge, e.g. about products, markets, and processes, with knowledge about data. This data knowledge is vital for organizations to be able to effectively utilize and sovereignly govern data in course of the digitization, particularly when industrial enterprises approach the implementation of Industry 4.0. As this is an area that academics have paid little attention to so far, the goal of the research presented in this paper is to propose a "data knowledge model" that conceptualizes data knowledge by identifying relevant model entities and their relationships. Various industry case studies and literature research have been conducted based on a DSR approach.

The evaluation of the resulting model proved its relevance and validity. From a scientific perspective it provides a basis for further research and proposes a common understanding for academic discourse. Practitioners profit from the consolidation of insights from different companies to a "best practice" model that presents a starting point that would have benefitted the case studies. However, the representation of tacit knowledge remains a concern demanding for a more detailed discussion.

The consideration of data as an asset proves that data are of equal importance for Industry 4.0 scenarios as physical assets. Given that data knowledge constitutes the content of a data asset's Administration Shell, we are able to contribute a partial model that covers various sections of the RAMI4.0. As a result, our data knowledge model proves to be relevant for I4.0 contexts. Nonetheless, new entities relevant for data knowledge in these environments require additional research.

Further efforts could be dedicated to exploring the capabilities required for data knowledge management and the supporting application architectures.

## References

[1] Bärenfänger, R., Otto, B.: Proposing a Capability Perspective on Digital Business Models. Presented at the 2015 IEEE 17th Conference on Business Informatics (CBI), Lisbon, Portugal July (2015).

[2] Hermann, M., Pentek, T., Otto, B.: Design Principles for Industrie 4.0 Scenarios. In: Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS-49 2016). , Kauai, HI, USA (2016).

[3] Yoo, Y., Henfridsson, O., Lyytinen, K.: Research Commentary—The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. Inf. Syst. Res. 21, 724–735 (2010).

[4] Bärenfänger, R., Otto, B., Österle, H.: Business value of in-memory technology – multiple-case study insights. Ind. Manag. Data Syst. 114, 1396–1414 (2014).

[5] Leveling, J., Edelbrock, M., Otto, B.: Big Data Analytics for Supply Chain Management. In: Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management (IEEM 2014). pp. 918–922. IEEE Computer Society, Selangor Darul Ehsan, Malaysia (2014).

[6] Timm, C., Weichert, F., Fiedler, D., Prasse, C., Muller, H., ten Hompel, M., Marwedel, P.: Decentralized Control of a Material Flow System Enabled by an Embedded Computer Vision System. In: 2011 IEEE International Conference on Communications Workshops (ICC). pp. 1–5. IEEE Computer Society, Kyoto, Japan (2011).

[7] Kagermann, H., Wahlster, W., Helbig, J.: Recommendations for implementing the strategic initiative INDUSTRIE 4.0. Final report of the Industrie 4.0 Working Group.

[8] Mithas, S., Ramasubbu, N., Sambamurthy, V.: How Information Management Capability Influences Firm Performance. MIS Q. 35, 237–256 (2011).

[9] Otto, B., Aier, S.: Business Models in the Data Economy: A Case Study from the Business Partner Data Domain. In: Alt, R. and Franczyk, B. (eds.) Proceedings of the 11th International Conference on Wirtschaftsinformatik (WI2013). pp. 475–489. , Leipzig, Germany (2013).

[10] Franklin, M., Halevy, A., Maier, D.: From Databases to Dataspaces: A New Abstraction for Information Management. SIGMOD Rec. 34, 27–33 (2005).

[11] Otto, B., Jürjens, J., Schon, J., Auer, S., Menz, N., Wenzel, S., Cirullies, J.: Industrial Data Space. Digital sovereignity over data. Fraunhofer-Gesellschaft, München (2016).

[12] Otto, B., Österle, H.: The Corporate Data League: One Approach for Cooperative Data Maintenance of Business Partner Data. In: Corporate Data Quality: Prerequisite for Successful Business Models. pp. 168–174. Springer Gabler, Berlin (2015).

[13] Henderson, D., Mosley, M., Brackett, M.H., Earley, S. eds: DAMA guide to the data management body of knowledge (DAMA-DMBOK guide). Technics Publications, Bradley Beach, N.J, USA (2009).

[14] Otto, B.: Enterprise-Wide Data Quality Management in Multinational Corporations, (2012).

[15] EFQM: EFQM Framework for Corporate Data Quality Management: Assessing the Organization's Data Quality Management Capabilities. , Brussels (2011).

[16] White, A.: 2013 Strategic Road Map for Enterprise Information Management. Gartner, Inc. (2013).

[17] Thomas, G.: The DGI Data Governance Framework. The Data Governance Institute, Orlando, FL, USA (2014).

[18] Khatri, V., Brown, C.V.: Designing data governance. Commun. ACM. 53, 148–152 (2010).

[19] IBM: IBM Data Governance Council Maturity Model. Building a roadmap for effective data governance. IBM Corporation (2007).

[20] Gold, A.H., Malhotra, A., Segars, A.H.: Knowledge Management: An Organizational Capabilities Perspective. J. Manag. Inf. Syst. 18, 185–214 (2001).
[21] Cohen, J.F., Olsen, K.: Knowledge management capabilities and firm performance: A test of universalistic, contingency and complementarity perspectives. Expert Syst. Appl. 42, 1178–1188 (2015).
[22] Nonaka, I., Takeuchi, H.: The knowledge-creating company: how Japanese companies create the dynamics of innovation. Oxford Univ. Press, New York, NY, USA (1995).
[23] Alavi, M., Leidner, D.E.: Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research, (2001).
[24] Denford, J.S.: Building knowledge: developing a knowledge-based dynamic capabilities typology. J. Knowl. Manag. 17, 175–194 (2013).
[25] Wallace, D.P., Fleet, C.V., Downs, L.J.: The research core of the knowledge management literature. Int. J. Inf. Manag. 31, 14–20 (2011).
[26] Nielsen, B.B., Michailova, S.: Knowledge Management Systems in Multinational Corporations: Typology and Transitional Dynamics. Long Range Plann. 40, 314–340 (2007).
[27] Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. J. Manag. Inf. Syst. 24, 45–77 (2007).
[28] March, S.T., Smith, G.F.: Design and natural science research on information technology. Decis. Support Syst. 15, 251–266 (1995).
[29] Österle, H., Otto, B.: Consortium Research. A Method for Researcher-Practitioner Collaboration in Design-Oriented IS Research. Bus. Inf. Syst. Eng. 52, 283–293 (2010).
[30] Yin, R.K.: Case Study Research: Design and Methods. Sage Publications Ltd., Los Angeles (2013).
[31] Susman, G.I., Evered, R.D.: An Assessment of the Scientific Merits of Action Research. Adm. Sci. Q. 23, 582–603 (1978).
[32] vom Brocke, J.: Design Principles for Reference Modeling: Reusing Information Models by Means of Aggregation, Specialisation, Instantiation, and Analogy. In: Fettke, P. and Loos, P. (eds.) Reference Modeling for Business Systems Analysis. pp. 47–76. IGI Global, Hershey, PA, USA (2007).
[33] Myers, M.D., Newman, M.: The qualitative interview in IS research: Examining the craft. Inf. Organ. 17, 2–26 (2007).
[34] Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In: Proceedings of the 17th European Conference on Information Systems (ECIS 2009). pp. 2206–2217. , Verona (2009).
[35] Holtkamp, B., Steinbuß, S., Gsell, H., Löffeler, T., Springer, U.: Towards a Logistics Cloud. Presented at the 2010 Sixth International Conference on Semantics, Knowledge and Grids November (2010).
[36] Böhmer, M., Schmidt, M., Weissenberg, N.: Seamless Interoperability in Logistics: Narrowing the Business-IT Gap by Logistics Business Objects. In: ten Hompel, M., Rehof, J., and Wolf, O. (eds.) Cloud Computing for Logistics. pp. 77–117. Springer, Berlin, Heidelberg (2015).
[37] Cooper, H.M.: Organizing knowledge syntheses: A taxonomy of literature reviews. Knowl. Soc. 1, 104–126 (1988).
[38] Rowley, J.: The wisdom hierarchy: representations of the DIKW hierarchy. J. Inf. Sci. 33, 163–180 (2007).
[39] Zins, C.: Conceptual approaches for defining data, information, and knowledge. J. Am. Soc. Inf. Sci. Technol. 58, 479–493 (2007).
[40] Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., Hoffmann, M.: Industry 4.0. Bus. Inf. Syst. Eng. 6, 239–242 (2014).
[41] DIN SPEC 91345: Reference Architecture Model Industrie 4.0 (RAMI4.0), (2016).
[42] VDI/VDE-GMA, ZVEI eds: Status Report. Reference Architecture Model Industrie 4.0 (RAMI4.0). , Düsseldorf (2015).
[43] BMWi (German Federal Ministry for Economic Affairs and Energy) ed: Structure of the Administration Shell. (2016).
[44] Otto, B., Abraham, R., Schlosser, S.: Toward a Taxonomy of the Data Resource in the Networked Industry. Presented at the Logistics in the Networked Industry : 7th International Scientific Symposium on Logistics, Cologne June 4 (2014).
[45] Hoberman, S.: Data modeling made simple: a practical guide for business and IT professionals. Technics Publications, Bradley Beach, N.J, USA (2009).
[46] Österle, H., Höning, F., Osl, P.: Methodenkern des Business Engineering. Ein Lehrbuch. (2011).
[47] The Open Group: TOGAF Version 9.1. (2011).
[48] The Essential Project: Essential Meta-Model Reference, http://www.enterprise-architecture.org/documentation/doc-meta-model. Accessed on 01 Mar 2016
[49] Mikloš, J.: A meta-model for the spatial capability architecture. J. Theor. Appl. Inf. Technol. 43, 301–305 (2012).
[50] Scheer, A.-W.: Architecture of integrated information systems : foundations of enterprise modelling. Springer, Berlin, Heidelberg (1992).
[51] Schmidt, A.: Entwicklung einer Methode zur Stammdatenintegration, (2010).
[52] Päivärinta, T., Tyrväinen, P., Ylimäki, T.: Defining Organizational Document Metadata: A Case Beyond Standards. In: ECIS 2002 Proceedings. , Gdansk, Poland (2002).