A METHOD FOR THE IDENTIFICATION AND DEFINITION OF INFORMATION OBJECTS

(Research-in-Progress)

Abstract: Due to the high complexity of their application and process landscapes a large number of today's companies, particularly multi-national enterprises acting on a global scale, are challenged by an increased lack of transparency with regard to their fundamental information objects. The paper at hand introduces a method that is intended to remedy this lack and ensure consistency among information objects on a conceptual level as a first vital step for designing an Enterprise Information Architecture. The scope of the method is not limited to a one-time effort, but rather constitutes an iterative approach for a continuous perpetuation and improvement of a consistent set of defined information objects. The whole method is designed for application in the context of large-scale companies that, due to their size and international scope, dispose of a certain complexity and inconsistency in regard to their information objects.

Key Words: Information Objects, Transparency, Corporate Data Dictionary, Information Architecture

INTRODUCTION

Motivation

In today's companies, particularly multi-national enterprises acting on a global scale, historically grown systems and application landscapes, as well as processes that are not harmonized and consistent, are nothing unusual. Problems arise when systems (and even employees) need to communicate with each other across their functional boundaries (formed e.g. by line-of-business, region etc.), i.e. when information is exchanged across processes and organizational units. Examples for such requirements are manifold, ranging from post-merger business integration to hybrid product strategies combining physical products and services from multiple lines of business. What is needed in situations like these is a common language and, hence, unambiguously and consistently defined entities that represent essential objects of a company's environment, such as vendors, customers, product hierarchies etc. [25]. Or, as [5] put it: "A key challenge of data quality is an incomplete or unclear set of semantic definitions of what the data is supposed to represent, in what form".

Transparency, as one of the pivotal terms of this article, describes the need for identifying the fundamental information entities an enterprise works with and for enhancing their common understanding. The term reflects the goal to give answers to the core questions regarding a company's Information Architecture, namely:

- Which information does a company work with?
- Who is responsible for the information?
- By which processes is the information used?
- Which applications access the information?
- How is the information exchanged between different applications?

The identification and description of essential information objects and their main characteristics (be it in a model or a glossary) constitutes a first important step towards answering these questions, particularly the first four of them. Therefore, the transparency on information objects is the basis for further analysis with

regard to the Information Architecture, most notably the application architecture and the information flows between applications, having direct influence on data quality and data integration issues [33]. The elaboration of a Corporate Data Dictionary (CDD) in which all information on these entities is stored is a critical contribution to achieving lasting improvements in the quality of data [10] by facilitating the optimal creation, use and sharing of information by the business. In contrast to traditional data dictionaries, a CDD is intended to store business metadata for which still little support is available [37] rather than technical metadata.

Although a number of frameworks exist in the field of enterprise architecture and metadata management that describe constituent artifacts, an organization-wide architected approach how to identify mission-critical information entities and describe them unambiguously is still missing. The Open Group Architecture Framework (TOGAF) for example specifies which descriptions, models and principles are needed for the transition from an as-is to a to-be architecture, however, there is no clarification how to proceed in order to obtain these artifacts as the procedure models is too generic [30].

The paper at hand provides a methodology that tries to fill this gap considering approved approaches and concepts in the research field. It shows how metadata repositories (such as a Corporate Data Dictionary) can be employed for increasing transparency on the company's information entities and presents a method – that we call METIO (Method for Establishing Transparency on Information Objects) – that enables an organization to successfully create and, most notably, keep transparency on and consistency among business relevant information objects with the help of metadata. The method we developed in our research intends to respond to the requirements to provide reliable and consistent information on a company-wide level and to assign dedicated roles for an organization's essential information entities [19].

Research Approach

The research follows the principles of design science which state that useful solutions must be obtained through the design and evaluation of models, methodologies and systems [42]. The particular focus lies on the construction of deployable artifacts, such as models and methods, that serve practitioners for solving real-world problems [20, 22]. In our case, the deployable artifacts consist of a method and a corresponding tool (the CDD) for storing the information object definitions that are elaborated with the help of the method. According to design science, our research provides both solutions relevant for defined business requirements (in terms of a method) and contributes to the advancement of the scientific body of knowledge by building on and extending established concepts [16].

The context of our research is set by an applied research project, which is being conducted in collaboration between a university and a number of industrial partner companies. Over a timeframe of two years, the consortium project focuses on the development of methods, tools and best practices to support Data Quality Management. Development of the proposed method METIO involved both university researchers and subject matter experts from the partner companies interacting tightly over the project duration. In the course of the collaboration, an action research approach was followed in order to provide the researchers with a detailed and continuous flow of information on the research subject and its context [8]. This necessitates a long-term involvement of the IS researcher who plays a helping role within the project by using his own scientific expertise to perform a diagnosis of the problem in the practitioner's setting and identify some appropriate actions [4, 20]. It accounts for the fact that complex social systems in which humans interact using information technologies can only be understood by inducing change and observing the effects of such intended change in a real-world organizational setting [3, 26]. Although both design science and action research presume intervention in a real world, we see the latter as being context-specific (e.g. in a certain enterprise) and utilize the iterative five step action research process recommended by [3] consisting of 1) Diagnosing, 2) Action Planning, 3) Action Taking, 4) Evaluating, and 5) Specifying Learning mainly for problem diagnosis and artifact evaluation. This process is accompanied by a design process in which the method and the tool are elaborated based on the findings of the action research studies.

The two real-life cases in which we conduct our action research are described in the section "Action

Research Studies". Their contribution to our research is twofold: Firstly, they represented the starting point for identifying the company-specific problems and requirements and, therefrom, deriving the procedure model of the method as the essential result artifact of our research. Secondly, the cases enable a real-world application of METIO, allowing for a comprehensive evaluation of the method and, consequently, significant findings for further improvement.

The remainder of the paper is structured as follows: the subsequent section provides the conceptual background for our research by firstly outlining a synoptic definition for metadata based on respective literature and secondly delineate the central terms used within this paper. Thereafter, we give an overview on the action research approach pursued and describe the projects which form the collaborative environment for the development of METIO in more detail. The following chapter contains the METIO methodology in more detail and, based on a peer literature review, derives how information objects can be unambiguously defined. The paper closes with a short conclusion and the outlook for further research.

BACKGROUND

Metadata

Metadata can be defined in general as data that describes other data (their meaning and properties) [7] demarcating them from other data types, such as transaction and master data. More precisely, we use the term to determine important characteristics that need to be known for either database and application engineering [13] or the general, semantically unambiguous understanding of data within the enterprise. Metadata, accordingly, facilitates the identification, retrieval, use and management of data as they allow an organization to better understand its data sources and definitions [24]. [40] summarizes their function by "seeing metadata as the means by which the structure and behavior of data is recorded, controlled, and published across an organization". The most comprehensive definition is provided by [23] who describes metadata as "all physical data (contained in software) and knowledge (contained in employees) from inside and outside an organization, including information about the physical data, technical and business processes, rules and constraints of the data, and structures of the data used by a corporation" [23]. The significance of this definition derives from its strong business orientation that we pursue in this paper as well. Herein, we utilize metadata in the form of attributes that need to be defined for specifying syntax and semantics of each information object.

Due to their high semantic content, metadata are the fundamental components for the design of information object models as well as Corporate Data Dictionaries, serving as an original source for the definition of data elements [9]. By maintaining information about the source of data, their (change) history or responsibility, metadata facilitates the challenge of keeping data consistent, accurate and complete. And high quality data, in turn, is pivotal for enabling service-oriented business applications [25], for helping to increase the validity of strategic decisions [36] and allowing high regulatory compliance [14]. Moreover, they enable a time- and cost-efficient way of retrieving, managing, evaluating and using appropriate information through precise queries which increases the confidence of users in data and augments the decision-making quality [23]. The semantic content is essentially provided by standard (textual) definitions of the according data entities. Metadata management denotes the assignment of these definitions to data as well as their maintenance in a centralized metadata repository, such as a CDD [12].

Definition and Classification of Basic Concepts

As stated in [19], the definition of an Information Architecture for improving the quality of management

decision making constitutes an essential process within IT governance and management. This includes among others an Enterprise Information Architecture Model in which the applications and information flows are being mapped as well as “the development of a corporate data dictionary” in order to facilitate a common understanding of data amongst IT and business users [19].
In the Corporate Data Dictionary we describe information objects with the help of metadata (see previous chapter). Referring to information objects, we would like to clearly demarcate the concept at this point of the paper from related terms, such as business objects and data objects. Within this paper we position business objects on a process level representing the input and output of business tasks, e.g. the entities that are exchanged within and between business processes. These business objects are relevant to business experts and generally described (if at all) in a simple textual form or an enumeration of their constituent attributes – similar to the business object description proposed by [32]. The definition contains a coarse-granular description of characteristics relevant from a business (process) perspective. Data objects on the other hand, are technical representations of these business objects on a system level. In most cases these entities are mapped in a more formalized way and contain more technical attributes, such as data types, field lengths etc. The process- and application-specific views on of both business and data objects respectively cause ambiguity and lead to a m:n relationship between the process-specific variants of a business object and the number of different data objects representing the business object on a system level. Addressing this problem, we include an additional level in between containing information objects that constitute business relevant entities on a logical level. This information object layer represents a layer of reference on which process-specific business objects and application-specific data objects are mapped. Information objects are described with their semantics as well as structure (consisting of relationships to other information objects) and go beyond purely business-oriented definitions. They are mapped and described with their entire set of attributes and consequently represent an integrated, cross-applicational view of both business and data object characteristics. By contrast, data objects are application-specific storing a subset of characteristics of the corresponding information object [34].

ACTION RESEARCH STUDIES

Automotive Manufacturer

Company and Problem Presentation

The first action research study refers to the Passenger Car Group division of a global automotive manufacturer with a turnover of € 99.4 billion in 2007 and more than 270,000 employees worldwide. The business factors with impact on data management mainly derive from the overall corporate profitability targets. They materialize in the need for common reporting structures to allow for comparison of different locations and in the constant requirement to reduce selling, general and administrative (SG&A) costs.
Among the reasons for current shortcomings in data management meeting the requirements is the complexity of the application architecture. In the past, application planning was accounted for on plant level, i.e. various business units were responsible for the task, leading to a total number of more than 2,000 applications in operation. Today, with integrated business processes that span multiple units, difficulties occur with mismatching definitions of information objects, unclear source systems for certain data objects, and numerous point-to-point connections between different application systems.
Apart from that, with continuing demands to reduce SG&A costs, the cost for IT has to be reduced as well. However, with the current lack of transparency regarding the relationship of information objects to business processes as well as application systems, consolidation of the system landscape is not an option

as it is not clear which systems hold redundant data or data critical to the business, and which systems serve which business processes.
As a response to that, part of the architecture planning initiative is an effort to create transparency regarding information objects, especially with regard to their relation to business objects used in business processes and to data objects held in application systems.

Pursued Approach

The high complexity of the company's application and data architecture lead to the awareness that a possible solution for increasing transparency should be constrained to a manageable pilot process in a first step. Consequently, in the initial Diagnosing phase we identified the Customer Order Process as the most appropriate for our goals as it satisfies the defined requirements, namely to span over multiple organizational units (purchasing, production, customer service etc.) and several information systems (as a lack of transparency becomes notably evident beyond functional boundaries), to possess limited complexity, and to be well defined. During the Action Planning phase we jointly worked out a first procedure model in expert discussions and determined work packages as well as relevant contact persons for necessary interviews.
We started the identification process of the relevant information objects (Action Taking) by analyzing the existing process documentation in close collaboration with Business Process Engineers who have a sound knowledge of the whole process across functional boundaries. This, firstly, enabled us to understand the entire process and, secondly, elaborate an initial list of process critical information objects with their assignment to corresponding process segments. Thereafter, we validated and detailed this first set of entities in several interview rounds with subject matter experts from the different business divisions involved in the Customer Order Process. In order to concentrate in the following definition process on the essential information objects only, we used the classification scheme provided by the Storage Networking Industry Association (SNIA) to prioritize the entities to be defined. The SNIA classification scheme distinguishes five categories (from "Not Important" to "Mission Critical") [39] that can be adopted for business data classification. Simultaneously, we worked out a metadata model[1] together with the project team including all relevant attributes for a comprehensive and unambiguous description of each identified information object. Based on this metamodel we prepared an interview guideline and an information object description template for the interviews that we conducted.
During the interviews with the different departments the different views on the identified entities were aggregated to a holistic information object description. Problems with differing or even contradictory understandings for some of the information objects were resolved in consolidation workshops with representatives from the corresponding business units. After finalization of these consolidation workshops – that are still under way – the specification of each information object will be stored in the company's architecture management tool whose structure for information object definition was adjusted to the metadata model presented in this paper.

Public Infrastructure Operator

Company and Problem Presentation

The second case refers to a European national railway network operator. The company – being part of an international transport and logistics group – is 100 percent state-owned; among its major business functions is the maintenance, repair and renewal of infrastructure assets such as tunnels, bridges and

[1] The metadata model as well as its derivation is described in more detail in one of the later sections.

tracks. With more than 42,000 employees the company generated an overall turnover of approximately € 4 billion in 2007.

There are two major business requirements with impact on the management of so-called infrastructure data, i.e. master data describing infrastructure assets. First, new regulations lead to a change of the monetary endowment from the public authorities for infrastructure maintaining purposes. In the past, the money was approved basically on a case-wise basis whereas nowadays a lump-sum strategy is established. The yearly amount is directly related to the quality of infrastructure data, i.e. on the ability of the company to provide accurate, consistent and timely information on the number and maintenance status of the national infrastructure assets. The second business driver is of operational nature. The majority of maintenance and construction activities require involvement of virtually all business functions such as construction planning, timetable-planning, asset management, maintenance etc. However, due the traditional line and staff organization, the business functions are poorly integrated leading to long cycle times, lots of manual rework, high process costs through double-work and a lack of transparency on the operational status of infrastructure assets. With regard to data management, the disintegration of business functions results in a lack of an integrated view on and responsibilities for information objects leading to ambiguous understandings of commonly used entities. Due to these challenges the company is undertaking an initiative to establish a common infrastructure data management aiming at the increase of transparency of the fundamental information objects and at efficiency gains in the data management domain.

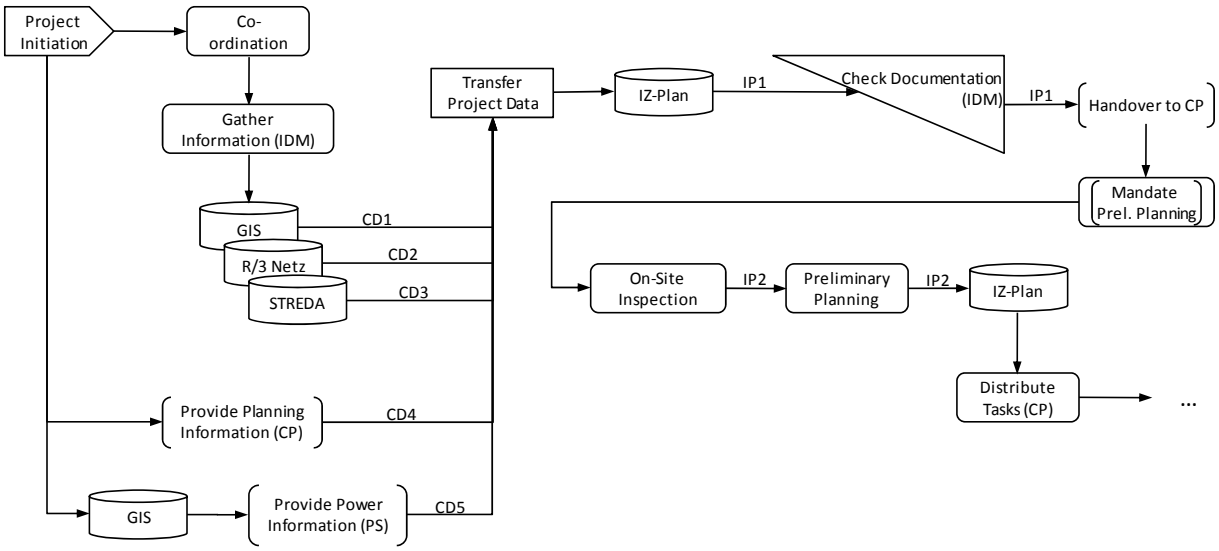
Pursued Approach

The identification process of the essential information objects pursued at the company was likewise as described in the first action research study. However, as a first learning of the experience made during the other project the information flow analysis on a process level was enhanced by an analog analysis of the application architecture in order to identify data objects that should ideally correspond to the identified business objects from the process level (Action Planning). This enabled us to get our attention on the discrepancy between the processes and the applications supporting them. Again, the analysis was deliberately constrained to one single process, in our case the reconstruction of a train platform.

In several workshop sessions with process owners and line managers from concerned departments during the Action Taking phase we were able to clearly describe the whole process as well as the business objects exchanged between and used within process segments using an information flow analysis. For mapping these information flows, we based our approach on the notation for designing Information Product Maps (IP-MAPs) which was proposed by [10, 35] for systematically mapping the manufacturing process of information products. An extract of the resulting information flow chart is depicted in Figure 1.

After having identified the applications that provide data for the different business process segments, the corresponding application owners began with the identification of the interfaces and data objects stored within their applications and exchanged between them based on existing system diagrams and data flow maps. Due to the relatively low number of identified information objects a prioritization was not considered necessary.

With regard to the description of the identified information objects we pursued an iterative process, based on an initial verbal description of the term to be defined, comparable to a simple glossary entry. Theses entries were stored and could be modified in a relatively simple way via the company's internal wiki. The rationale behind this approach was to facilitate employees to get involved in the definition process and therefore, increase contribution from different departments that might have diverging understandings of an information object. Thereafter, based on these verbal definitions a common definition could be derived in a concerted process and the necessary attributes for each entity according to the metadata model could be specified leading to a comprehensive Corporate Data Dictionary.



CD1: Railroad Network Data
Geographical Coordinates
Mast
Signal Flow
Signaling Equipment
Signal Aspect
Platform Edge
Topology
Route
Cable Stranding
Power Supply

CD2: Maintenance and Repair
Roadbed and Track
Railroad Embankment
Buffer Stop
Standard Cross-Section
Layer Composition & Material
Rail and Rail Fastening
Rail-Support Element
Turnout System
Control Equipment

CD3: Route Data
Regulation Framework
Stop Station
Characteristics (E, Sig)
Narrow Points

CD4: Route Planning
Real Estate Plans/Certificates

CD5: Conduit
Distribution Rack
House Connection
Wiring

Information Product (IP):
IP1 ... Infrastructure Documentation
IP2 ... Construction Plan

Organizational Units:
IDM ... Infrastructure Data Management
CP ... Construction Planning
PS ... Power Supply

Figure 1. Information flow diagram for the process "Reconstruction of Train Platform" (extract)

DESIGN OF A METHOD FOR ESTABLISHING TRANSPARENCY ON INFORMATION OBJECTS

The action research studies described in the previous section provided essential insights for our goal to define a methodological approach for identifying and unambiguously describing a company's essential information objects. These insights were further enriched by an in-depth literature research on established methods for system analysis and engineering. As a result, we integrated both theoretical findings [2, 38] and best practices from solution providers in adjacent fields. The entire specification of the method incorporates the definition of all constituents of a method according to the Method Engineering [15, 21], among others a precise description of each activity, the responsible roles[2] and techniques for achieving the required results that has been left out in this paper due to space limitations.

From the two action research studies described in the previous section we derived the following requirements for our method:

- Distinction between information object identification and definition;
- Combined top down and bottom up approach, comprising the analysis of entities on a process level (business perspective) and a system level (application perspective) in order to guarantee consistency between these layers and a common understanding of data amongst business and IT users;

[2] A shortened presentation of the roles necessary for our method can be found in the next section.

- Both identification and definition of the information objects are highly iterative processes necessitating several rounds of interviews and workshops to be conducted;
- Iterative approach with regard to the scope of application (from a single process to company wide application);
- Appropriate tool for storing and maintaining information object definitions.

Procedure Model

Based on these findings, and backed by the theoretical knowledge gained from desk research, Figure 2 illustrates the overall procedure model for establishing company-wide transparency on fundamental information objects.

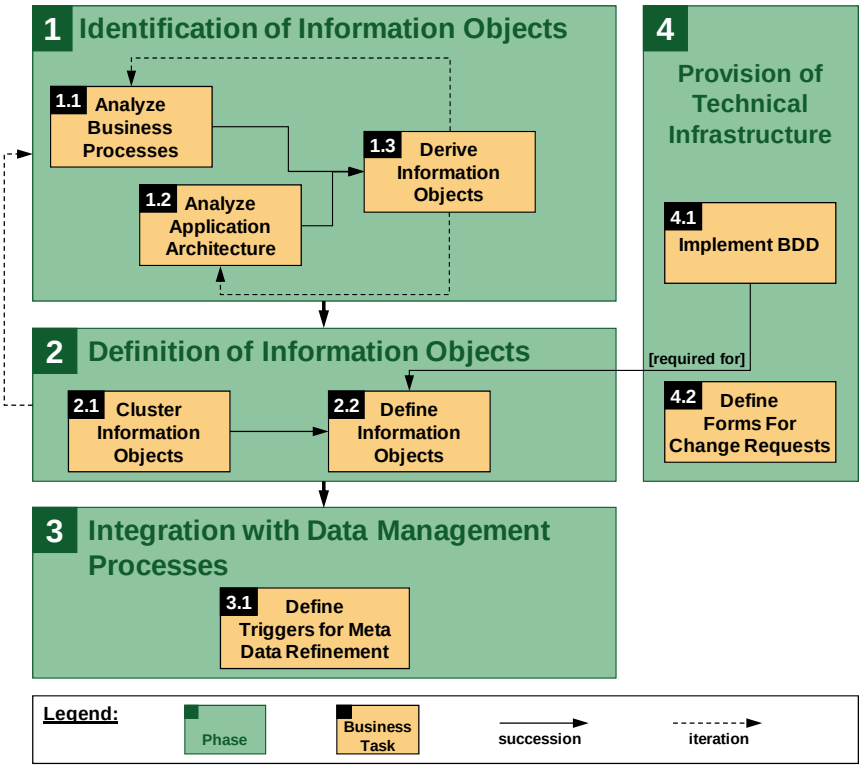


Figure 2. Overall procedure model

The roles that are used within the METIO method (and in the following as well) are adopted from already established roles in the field of Data Governance [6, 12, 13, 41], that defines roles and implements corporate-wide accountabilities for Data Quality Management. Three of them are of particular interest within METIO and are therefore briefly introduced [41]:

- Corporate Data Steward (CDS) who puts data quality relevant decisions into practice, enforces the adoption of standards, helps establish data quality metrics, standards and targets;
- Business Data Steward (BDS) who details corporate wide data quality standards and policies for his or her area of responsibility from a business perspective;
- Technical Data Steward (TDS) who provides standardized data element definitions and formats, profiles source system details and data flows between systems.

The responsibilities and tasks have been adopted as far as possible and extended by aspects specific to the definition and harmonization of information objects [24], as Metadata Management is considered as a pivotal task of Data Governance.

The process starts with the identification of relevant information objects. This first step has to be executed

in a combined top down and bottom up approach that allows for integrating essential entities both from a process and a system perspective. The most substantial entities derived from these analysis tasks are either data objects without an equivalent business object on a process level, or, the other way around, business object with no analogue data object in a company's IT systems. Moreover, business objects with multiple representations on the system level constitute further entities relevant for consolidation. For the purpose of a revelation of these misfits, a consciously separated execution of these two business tasks is recommended. This two-sided approach aims at ensuring a common understanding of terms and data amongst IT and business users [19].

For the analysis on a process level, the already existing process documentation (particularly process models) needs to be worked through by the Business Data Steward who needs to possess a sufficient understanding of the business process. If the output resulting from or being exchanged between business processes or process steps, such as business documents or goods, is mapped, these entities constitute candidates for possible business objects. In case of insufficient process documentation (additional) interviews with Business Process Owners and Engineers are necessary to obtain the missing information and essential business objects.

Simultaneously, from a system perspective and complementary to the identified business objects essential data objects are to be identified by the Technical Data Steward in collaboration with Application Owners and Data Engineers. As we learned in our action research studies, companies, particularly multinational enterprises, dispose of a variety of different systems and applications, real-world objects are often represented in a non-consistent way. This leads to a multitude of synonyms and homonyms inhibiting transparency and consistency. Consequently, it is necessary to consolidate the variety and obtain a redundant free mapping of the data objects on a system level. Therefore, application-specific data models as well as interviews with respective Data and Application Owners are helpful information sources.

The first phase concludes with a joint consolidation of the relevant entities resulting from the first two business tasks that needs to be executed by all Data Stewards in order to derive a consolidated and non-redundant set of information objects for definition. Hence, possible misfits between the process perspective (business objects) and the application perspective (data objects) need to be resolved. Within this business task further business or data objects might be detected, necessitating a new iteration of the before-mentioned tasks.

Thereafter, the actual definitions of the essential entities are provided. Here different definition process variants depending on the roles involved in the definition process are possible. The predefined process variants as well as the executing roles are assigned in a first step to each of the information objects. Each of the information objects can be specified according to the process variant it is assigned to. The sequence then encompasses the following activities (with corresponding responsibilities):

- Provide terminological and contextual metadata on information objects (Business Data Steward),
- Define relationships to and dependencies from other information objects, assign ownership and provide administrative metadata (Corporate Data Steward),
- Specify configuration requirements (Technical Data Steward),
- Define security level (optional) (Data Security Agent),
- Add compliance-relevant metadata (Legal Department).

When the definition of an information object is finalized, an intensive review of the results by the CDS or a central board specifically set up for this purpose is necessary in order to ensure completeness of the specification as well as consistency with other definitions. This emphasizes the need for governance that emerges when information objects are to be defined on an organization-wide level. The control is needed to reconcile terms cross-functionally with other groups in the organization who may have a different usage of a certain term [28].

As noticed in both action research studies and expressed by the Data Stewards in the companies we worked with, a major challenge remains the retention of the once established transparency and to keep a good quality of the defined information objects. This is particularly true in the dynamic environment in which companies operate nowadays: new products are launched, regulations change, mergers and

acquisitions lead to new business vocabularies. And when business changes, this may lead to definitions which have been correct at one point in time but become obsolete over time. Hence, enterprises need to be able to change and adapt the definitions of relevant information objects or add new ones in the most flexible way possible [28]. This includes the possibility to issue change requests easily by the people using these information objects and to make sure that possible conflicts are resolved and the wording of the definition is kept accurate. The change request should be formalized by an appropriate "Request for Change" form and process as for example described in [29]. Therefore, we included process step 3.1 which is intended to guarantee the embedding of the metadata definition and maintenance process with the operational (meta-) data usage processes once the initial definition process is finalized.
We consider the process of establishing and maintaining a maximal transparency with the help of metadata as a nested and iterative process within the regular processes on a data level. This means that the triggers for transition to and from the metadata definition process need to be clearly defined. Therefore, it needs to be ensured that the preconditions for re-entering the metadata definition process in order to adapt and refine existing information object definitions, or integrate new ones, are regularly checked.

Metadata for Defining Information Objects

For the IO definitions in our CDD prototype a number of attributes can be maintained in order to allow for a comprehensive and unambiguous specification with a maximum of semantic information. Regarding this set of attributes, the question arises why we included exactly these metadata elements (and not others as well) and whether they actually allow us a sufficient and unambiguous definition of each information object. Therefore, we conducted an in-depth literature analysis including metadata standards in adjacent research fields such as computer as well as library and information science where metadata play an important role. These potentially relevant attributes were then discussed with domain experts and complemented with further characteristics that are important from a business perspective.
[31] identified a set of metadata elements as a result of their study of 19 contemporary public standards and specifications for document management that were considered potentially relevant. From the entirety of stated metadata elements (i.e. attributes) the authors extracted the ones stated most often in the standards and refined them by adding attributes from organizational needs obtained in discussion with representatives from the domain. Most of the 14 metadata elements, so-called "core elements", derive from the Dublin Core Metadata Specification [11], the Australian Government Locator Service [1] and the ISO/IEC 11179-3 and -4 (specification and standardization of data elements and formulation of data definitions) [17, 18]. The identified metadata elements are summarized in alphabetical order in Table 1. Besides, we included a short description and their equivalents from our CDD.

Element name	Description	CDD attribute
Data type	Data type of a metadata element (e.g. character string)	Data Type and Field Length
Default value	Default value of a metadata element	---
Definition	Short description of a metadata element; what is the content of the element	Definition
Example	Examples of the values assigned to a metadata element	Potential Values
Identifier	Unique identifier of a metadata element	Provided by an unambiguous name
Max. occurrence	Number of values assigned to a metadata element. The repeatability of the metadata element.	---
Name	Name of the metadata element	Name
Obligation	Obligation of a metadata element: mandatory (M), conditional (C) or optional (O)	---
Producer(s)	Organization/department/team/person/role, that produces the content of a metadata element and is responsible for it	Responsible Business Data Steward
Purpose and	Justification; why is this metadata element needed? How is it	o Rationale

comments	used? Other comments or instructions.	o Comment
Standard	Standard or specification, which defines the metadata element in question (name of standard and element).	---
Sub-elements	Sub-elements of a metadata element	Related Terms / Relationship
User(s)	Organization/department/team/person/role, that uses a metadata element	Validity Scope
Value qualifier	Name of the set of values or list of values that can be assigned to a metadata element. There can be one or more sets of values	Potential Values

Table 1: Attributes to describe metadata according to [31] and their equivalent CDD attributes

As Table 1 indicates, we used the majority of the identified attributes for our CDD, however, we adapted some of the elements with regard to their name and scope. The attribute "Purpose and comments" was split into two separate items and the first renamed into "Rationale" as this represents the underlying semantics more adequately. Attributes, such as "Max. occurrence" or "Default value", were omitted due to their minor relevance for our cases.

A similar synthesis was conducted by O'NEIL for the components necessary to provide a sound definition within a glossary [27]. From this list of attributes we identified a number of further components that we could use for the CDD in addition to the ones stated above (such as "Name" and "Examples"). The attributes appended are:

- "Related", "Narrower" and "Broader Term" were subsumed under the generalized / aggregated term "Relationship to other IO" that – in our case – incorporates the relationship to other information objects and can be a "is-a" (corresponding to a broader-narrower-term relationship) as well as a "see also" relationship;
- "Source" was slightly changed in its naming (to "Strategic Source") and meaning, signifying the original source of the information object rather than the source where the definition came from; and
- "Distinguishing Characteristics" and "Synonyms" were directly transferred to our own CDD model with the definitions as stated in [27].

The element "Replaced by" was conceptualized broader and realized in a slightly different way. It constitutes a possible value within the CDD attribute "Status" (with Draft, Final, Approved and Retired being other possible status values). In case of a "Replaced" status of an information object a relationship "replaces/replaced by" has to be assigned to the attribute "Related Terms" in order to correctly map the replacement of one entity with another.

As those studies clearly lack a specific business and implementation focus, the results could not be transferred directly to our CDD and had to be either adapted to our specific needs (as outlined above) or supplemented by further attributes. For this purpose we integrated the information gathered from interviews and discussions with domain experts from our research project to allow an implementation that serves the requirements of our project partners. This enabled us to complement the results of the literature review with their tacit experience and knowledge of the business context.

The elements added as a result of these interviews are either relevant for implementation (such as "Security Classification" referring to the security level etc.) or provide information for the embedding in a specific business context (such as "Subject Area", "Validity Scope" (of application within the organization), "Coding and Descriptive Conventions"). The relevance of attributes addressing security classification and encoding descriptions is also reflected in the metadata standards comparison by [7]. As information has to be considered within the context of the processes and applications in which they are used, the corresponding information can be maintained in the CDD under the attributes "Usage in Processes" and "Usage in Applications". Particularly the latter is needed within the scope of application architecture planning and development when certain applications are to be replaced or deprecated.

Lastly, we added three attributes that specify how each information object is maintained ("Maintenance

Procedure" and Maintenance Process Documentation") and instantiated ("Instancing Process Documentation") in order to help to keep the transparency and consistency on a constantly high level.
The attributes were clustered into categories based on the type of metadata. Figure 3 summarizes the attributes that need to be defined for unambiguously describing information objects according to METIO and illustrates the meaning of some of the attributes by the exemplary information object "Customer".

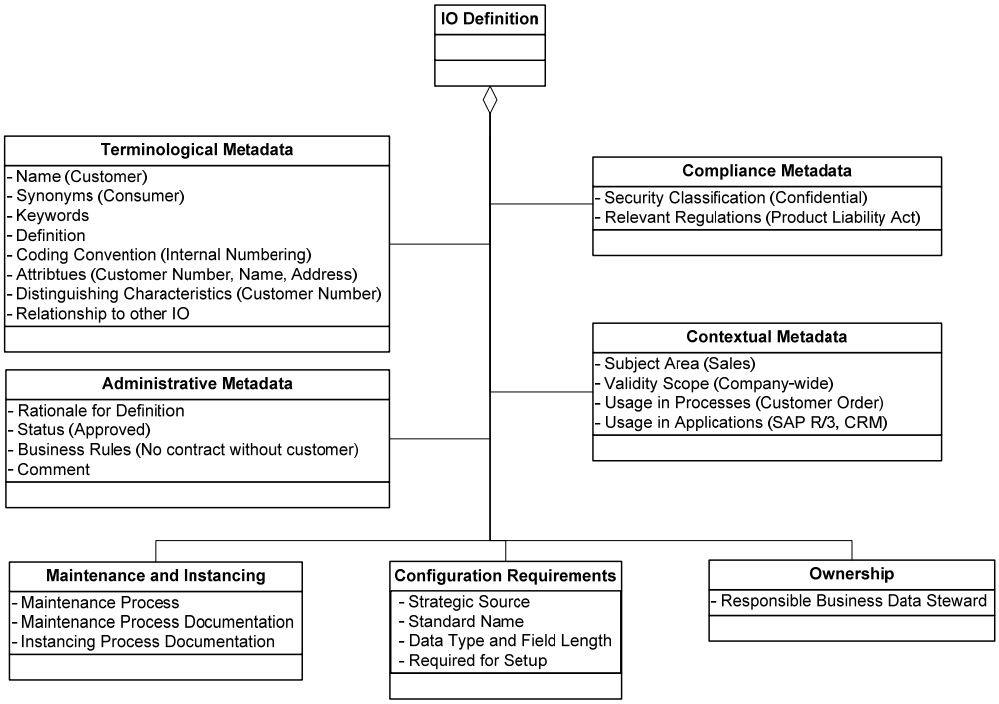


Figure 3. Attributes to be defined for comprehensive information object description in a CDD

Two attributes of the CDD metamodel are of particular importance. Firstly, the attribute "Distinguishing Characteristics" offers the possibility to include pertinent characteristics with specific values for each information object being defined. The attribute should not serve as a container for all existing properties but rather those characteristics that are specific to the information object being defined. This allows demarcating entities of the CDD more clearly from each other. Secondly, the exact characterization of the dependencies to other information objects is obtained by the attributes "Synonyms" (i.e. information objects with similar semantics) and "Related Terms". The latter is used to precisely describe the relationship to associated entities in order to provide structural information. Consequently, these attributes realize the added value of our CDD in comparison to simple glossaries.

CONCLUSION

The paper at hand introduced a method that is intended to enable companies to increase the transparency and consistency among information objects on a conceptual level. The method was developed using an action research approach based on two action research studies and an extensive literature research and is complemented by a corresponding tool for metadata storage and maintenance (the Corporate Data Dictionary). The method describes how relevant entities can be identified and then precisely defined. The (positive) consequences of unambiguously defined information objects are manifold: Firstly, they ensure a common understanding of important information objects for all entity users increasing significantly their productivity of work due to decreased search times or incorrectly stored data. Secondly, they

directly increase data quality within an enterprise as all instances of used information objects are specified according to a uniform and consent definition. This, thirdly, leads to a facilitated communication with people speaking literally the same language, and helps the business make more accurate decisions [27]. And lastly, they are a prerequisite for seamless semantic integration of heterogeneous applications that need to exchange data [19].

In the action research process depicted at the beginning of this paper, we are currently at the end of the third step (Action Taking) of the research process, finalizing the Corporate Data Dictionaries according to the METIO method in both real-life scenarios. The execution, as well as the Diagnosis and the Action Planning before, were carried out in close collaboration with practitioners from the corresponding companies and described in this article. After finalizing the adoption of our method within both organizations, it will be essential to evaluate the results of the implemented metadata repository and the metadata management processes, which represents an integral part of our future research. In order to assess and prove possible benefits on data quality we need to elaborate a metric for quantifiable evaluation of METIO. For this purpose and for the sake of further refinement, the adoption of the method in further real-word settings is intended.

Moreover, for further research, we consider the following issues as potential areas of interest:

- Utilization of metadata for data quality measurement,
- Metadata integration in other applications that are used by the ordinary employee in order to facilitate his work and improve the quality of his results,
- Extension towards an ontology-based, company-wide semantic web that allows for semantically enriched, intelligent search and real knowledge management.

REFERENCES

[1] Australian Government Locator Service "AGLS Metadata Element Set – Part1: Reference Description. Version 1.3." National Archives of Australia, Canberra, Australia 2002.

[2] Balzert, H. *Lehrbuch der Software-Technik - Software-Entwicklung*. Spektrum Akademischer Verlag. Heidelberg, 2000.

[3] Baskerville, R. L., Pries-Heje, J. "Grounded action research: a method for understanding IT in practice." *Accounting Management And Information Technologies,* 9(1999. pp. 1-23.

[4] Baskerville, R. L., Wood-Harper, A. T. "Diversity in information systems action research methods." *European Journal of Information Systems,* 7(1998. pp. 90-107.

[5] Berson, A., Dubov, L. *Master Data Management and Customer Data Integration for a Global Enterprise*. McGraw-Hill, 2007.

[6] Bitterer, A., Newman, D. "Organizing for Data Quality." Gartner Research, Stamford, CT 2007.

[7] Burnett, K., Ng, K. B., Park, S. "A Comparison of the Two Traditions of Metadata Development." *Journal of the American Society for Information Science*, 50(13). 1999. pp. 1209-1217.

[8] Checkland, P., Holwell, S. "Action Research: Its Nature and Validity." *Systemic Practice and Action Research,* 11(1). 1998. pp. 9-21.

[9] Chisholm, M. "Master Data versus Reference Data." *DM Review*, 16(4). 2006.

[10] Davidson, B., Lee, Y. W., Wang, R. Y. "Developing data production maps: meeting patient discharge data submission requirements." *Int. J. Healthcare Technology and Management*, 6(2). 2004. pp. 223-240.
[11] Dublin Core Metadata Initiative *Dublin Core Metadata Element Set, Version 1.1*. Dublin Core Metadata Initiative. 2008. Available at: http://dublincore.org/documents/dces.
[12] Dyché, J., Levy, E. *Customer Data Integration*. John Wiley & Sons. Hoboken, New Jersey, 2006.
[13] English, L. P. *Improving Data Warehouse and Business Information Quality*. 1th ed. John Wiley & Sons, Inc. New York, NY, 1999.
[14] Friedman, T. "Gartner Study on Data Quality Shows That IT Still Bears the Burden." Stamford: Gartner Group, 2006.
[15] Gutzwiller, T. *Das CC RIM-Referenzmodell für den Entwurf von betrieblichen, transaktionsorientierten Informationssystemen*. Physica. Heidelberg, 1994.
[16] Hevner, A. R., March, S. T., Park, J., Ram, S. "Design Science in Information Systems Research." *Management Information Systems Quarterly*, 28(1). 2004. pp. 75-105.
[17] ISO "Information Technology – Metadata Registries (MDR). Part 4: Formulation of Data Definitions." International Organization for Standardization (ISO), Geneva, Switzerland 1995.
[18] ISO "Information Technology – Metadata Registries (MDR). Part 3: Registry Metamodel and Basic Attributes." International Organization for Standardization (ISO) 2003.
[19] IT Governance Institute "CobiT 4.0: Control Objectives, Management Guidelines, Maturity Models." IT Governance Institute, Rolling Meadows/IL 2005.
[20] Lee, A. S. "Action as an artifact." in *Information Systems Action Research An Applied View of Emerging Concepts and Methods*, Ned Kock, Ed.: Springer US, 2007, pp. 43-60.
[21] Leist, S., Zellner, G. "Evaluation of Current Architecture Frameworks." in *Proceedings of the 21st Annual ACM Symposium on Applied Computing (SAC2006), April 23-27, 2006*, New York, 2006, pp. 546-1553.
[22] March, S. T., Smith, G. F. "Design and natural science research on information technology." *Decision Support Systems*, 15(4). 1995. pp. 251-266.
[23] Marco, D. *Building and Managing the Meta Data Repository. A Full Lifecycle Guide*. John Wiley & Sons. New Jersey, 2000.
[24] Marco, D., Smith, A. M. "Metadata Management & Enterprise Architecture: Understanding Data Governance and Stewardship." *DM Review*, 16(9-11). 2006.
[25] Newman, D., Logan, D. "Achieving Agility: How Enterprise Information Management Overcomes Information Silos." Gartner Research, Stamford, CT 2006.
[26] Nilsson, A. "Management of Technochange in an Interorganizational E-government Project." in *Proceedings of the 41st Hawaii International International Conference on Systems Science (HICSS-41 2008)*, 2008.
[27] O'Neil, B. K. *Business Metadata: How To Write Definitions*. Seiner, Robert S. 2005. Available at: http://www.tdan.com/i032fe01.htm.
[28] O'Neil, B. K. *Launching a Corporate Glossary*. 2005. Available at: http://www.b-eye-network.com/view/1014.
[29] OGC *ITIL - Service Transition*. TSO, 2007.
[30] Open Group *The Open Group Architecture Framework TOGAF - 2007 Edition (Incorporating 8.1.1)*. Van Haren. Zaltbommel, 2007.
[31] Päivärinta, T., Tyrväinen, P., Ylimäki, T. "Defining Organizational Document Metadata:

A Case Beyond Standards." in *Proceedings of 10th European Conference on Information Systems (ECIS)*, Gdansk (Poland), 2002.

[32] Scheer, A.-W. *ARIS - Modellierungsmethoden, Metamodelle, Anwendungen*. Springer-Verlag. Berlin et al., 2001.

[33] Schreiber, Z. *Semantic Information Architecture: Creating Value by Understanding Data*. dm review. 2003. Available at: http://www.dmreview.com/article.

[34] Schwinn, A. "Entwicklung einer Methode zur Gestaltung von Integrationsarchitekturen für Informationssysteme." Bamberg: Universität St. Gallen, Difo Druck, 2005.

[35] Shankaranarayan, G., Wang, R. Y., Ziad, M. "IP-MAP: Representing the Manufacture of an Information Product." in *Proceedings of the 2000 Conference on Information Quality*, Boston, 2000.

[36] Shankaranarayan, G., Ziad, M., Wang, R. Y. "Managing Data Quality in Dynamic Decision Environments: An Information Product Approach." *Journal of Database Management*, 14(4). 2003. pp. 14-32.

[37] Shankaranarayanan, G., Even, A. "The Metadata Enigma." *Communications of the ACM*, 49(2). 2006. pp. 88-94.

[38] Sommerville, I. *Software Engineering*. 8th ed. Pearson Studium, 2007.

[39] Thome, G., Sollbach, W. *Grundlagen und Modelle des Information Lifecycle Management*. Springer. Berlin, 2007.

[40] Tozer, G. *Metadata Management*. Artech House computing library. Norwood, Massachusetts, 1999.

[41] Wende, K. "A Model for Data Governance – Organising Accountabilities for Data Quality Management." in *Proceedings of 18th Australasian Conference on Information Systems*, Toowoomba, Australia, 2007, pp. 417-425.

[42] Wilde, T., Hess, T. "Forschungsmethoden der Wirtschaftsinformatik. Eine empirische Untersuchung." *Wirtschaftsinformatik*, 49(4). 2007. pp. 280-287.