A Method for Establishing Transparency on Integration Objects

Alexander Schmidt, Philipp Osl

Report No.:	BE HSG/ CC CDQ/ 2
Chair:	Prof. Dr. H. Österle
Version:	1.0
Date:	September 23, 2008

University of St. Gallen - for Business Administration, Economics, Law and Social Sciences (HSG)

Institute of Information Management
Müller-Friedberg-Strasse 8
CH-9000 St. Gallen
Switzerland
Tel.: ++41 / 71 / 224 2420
Fax: ++41 / 71 / 224 2777

Prof. Dr. A. Back
Prof. Dr. W. Brenner (managing)
Prof. Dr. H. Österle
Prof. Dr. R. Winter

Content

List of Abbreviations

BDD	Business Data Dictionary
BDS	Business Data Steward
BO	Business Object
CC	Competence Center
CDQ	Corporate Data Quality
CDS	Chief Data Steward
DB	Deutsche Bahn
e.g.	for example
ETIM	Elektrotechnisches Informationsmodell
i.e.	that is
IO	Integration Object
IT	Information Technology
IWI	Institut für Wirtschaftsinformatik
MDA	Model Driven Architecture
METIO	Method for Establishing Transparency on Integration Objects
OMG	Object Management Group
OWL	Web Ontology Language
RDF	Resource Description Framework
TDS	Technical Data Steward
UML	Unified Modelling Language
UNSPSC	United Nations Standard Products and Service Code

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 business objects. What is needed is a common language and, hence, unambiguously and consistently defined entities that represent essential objects of a company's environment.

The working paper at hand introduces a method that is intended to enable companies to increase the transparency and consistency among integration objects on a conceptual level. We present the Method for Establishing Transparency on Integration Objects (METIO) with all constituents defined by the Method Engineering approach: its metamodel, its roles, its procedure model (including the activities necessary to be carried out), the documentation model and the techniques. The scope of METIO 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 integration objects. The whole method is designed for being used in the context of large-scale companies that, due to their size and international scope, dispose of a certain complexity and inconsistency with regard to their integration objects.

1 Introduction and 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, i.e. when information is exchanged across processes and organizational units. What is needed is a common language and, hence, unambiguously and consistently defined entities that represent essential objects of a company's environment defining. "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" [Berson/Dubov 2007, p. 113].

The consequences of information that is or can be misunderstood may be the malfunction of business processes, applications and whole organizations. In its worst case, mission-critical business decisions can be based on wrong information because of false data interpretation. Examples of disastrous implications are manifold, be it the destruction of the Mars Climate Orbiter due to the usage of different metrics [O'Neil 2005a, p. 1], the death of patients because of prescription mistakes and mislabelled blood samples or the additional costs of the United States Postal Service resulting from undeliverable mail [cf. Pierce 2005]. These examples illustrate the importance of information as a company asset which should be managed actively based on an organization-wide, architected approach. Data must be shareable in a transparent fashion across the organization and be under corporate ownership, rather than serving personal interests [Tozer 1999, p. xx].

The concept of (business) metadata provides a promising approach to reduce these misunderstandings by adding context to data [O'Neil 2005a, p. 1]. It aims at making meaning explicit and providing definitions to important business terms and objects, data elements and abbreviations. However, this approach raises new questions: How do we determine which objects to focus on? How do I store and manage metadata? Are there appropriate tools[1]? What roles do I need to define in order to embed the concept within my organization and who "owns" the metadata? How do I integrate implicit business knowledge into metadata

[1] The term tool in this sense (and in the rest of the working paper) does not necessarily refer to a software tool but rather a mean that helps managing metadata (such as a metadata repository).

definitions? How do I guarantee that everybody has constant access to these metadata in their operational work?

Due to the little in-depth research on metadata and, more specifically, the assessment of the value of metadata, companies still fail to see its value [Shankaranarayanan/Even 2006, p. 88]. Most of the concepts dealing with metadata are rather vague and lack a guided procedure, particularly for keeping data quality high. Moreover, they are usually not aligned with business needs and the actual users (due to an overload of technical information) which may lead to inappropriate use [Olson 2003, p. 19].

The working paper at hand provides a method that addresses this topic and gives answers to the questions mentioned above. It shows how metadata repositories (such as Business Data Dictionaries) can be employed for improving data quality, and it presents a method – called METIO (Method for Establishing Transparency on Integration Objects) – that enables an organizations to successfully create and, most notably, keep transparency of and consistency among business relevant objects with the help of metadata.

Transparency, as the pivotal term and aim of METIO, describes the need for identifying the fundamental 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 infomation?
- By which processes is the information used?
- Which applications access the information?
- How is the information exchanged between different applications?

As stated in CobiT [IT Governance Institute 2005], 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 [IT Governance

Institute 2005, pp. 33f.]. The identification and description of essential business objects (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 business and data 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 [Schreiber 2003].

The rest of the working paper is structured as follows: Section 2 provides the conceptual foundation for our research by examining the elements necessary to describe a comprehensive method in our research field. Thereafter, we provide a synoptic definition for metadata based on respective literature and compare different tools regarding their appropriateness for business oriented metadata management. The following chapter describe the METIO method in more detail starting with its metamodel (Chapter 3.1) that maps all relevant design objects of the method (these objects are then precisely defined in the corresponding Glossary). While Chapter 3.2 describes all necessary roles, Chapter 3.3 specifies the procedure model with all activities, assigned roles and expected outputs (of each activity) for defining business objects unambiguously followed by the complete description of the documentation model and the necessary techniques in Chapters 3.4 and 3.5. Thereafter, Chapter 4 is dedicated to the description of METIO in a real-word scenario in one of our project partner companies. The working paper closes with a short conclusion and the outlook for further research.

2 Conceptual Foundation

2.1 Method Engineering

The term Method Engineering describes a systematic and structured process for the development, modification, and adaptation of methods by specifying method components and their relationships to each other (in terms of their semantics as well as with regard to time aspects) [Heym 1993, p. 61, Nuseibeh et al. 1996, p. 267]. The goal is to apply the principles of design methods to the field of applied computer science in order to allow for an engineering like design of information systems and related artefacts. BRINKKEMPER defines the term “method” as “an approach to perform a systems development project, based on a specific way of thinking, consisting of directions and rules, structured in a systematic way in development activities with corresponding development products” [Brinkkemper 1996, pp. 275 f.]. The structure of the method developed in the working paper at hand is based on the principles of method engineering as outlined by GUTZWILLER [cf. Gutzwiller 1994]. Based on a review of different approaches to method construction and implementation HEYM [cf. Heym 1993] and GUTZWILLER [cf. Gutzwiller 1994] identified five constituent elements of a method that are illustrated in a metamodel (see Figure 2-1) and used for the description of generic methods.

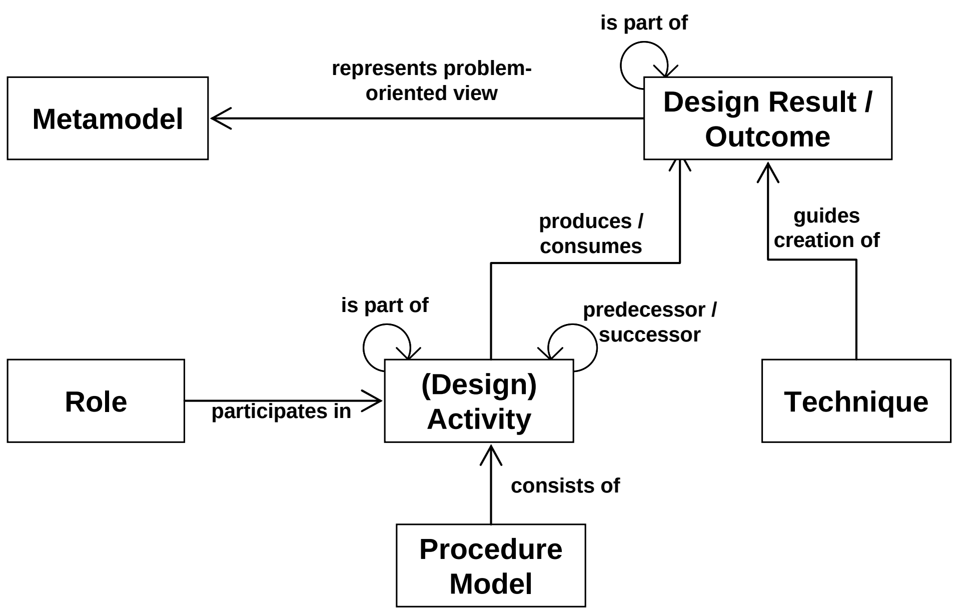

Figure 2-1: Method Engineering Metamodel (based on [Gutzwiller 1994, p. 13])

According to this metamodel, the *(design) activity* constitutes an integral part for the construction of a method. It represents a functional unit of action aimed at creating one or more defined outputs. Activities, on the one hand, can consist of consequent sub-activities and, hence, be hierarchically structured, on the other hand, a number of preceding and subsequent activities can compose a succession of activities [Gutzwiller 1994, p. 13]. The total sequence of activities of a method is mapped in a *procedure model*, showing a set of procedural guidelines which state when and in what order certain outputs need to be generated [Leist/Zellner 2006, p. 1548].

As already mentioned, activities generate or change a certain *outcome* (also referred to as *design result*) that in turn can be used as input by subsequent activities. Most frequently, design results are produced in the form of documents, however, other artefacts, such as models or products, are possible. The relationships and connections between different outcome artefacts are modelled in the *metamodel*, often referred to as information or metadata model of a method. It depicts the conceptual data model of the generated output [Gutzwiller 1994, p. 14]. In accordance with other authors [Goldkuhl et al. 1997, Greiffenberg 2003] the metamodel and the procedure model can be considered as pivotal constituents of a method.

Finally, *techniques* serve as guidelines used for instructing the method user how to create an output. In contrast to the procedure model, that describes the process in general (which design results are generated at which point of time), techniques characterize the procedure to perform a development activity in more detail [Brinkkemper 1996, p. 276]. Activities are carried out by single persons and/or organizational units in different *roles* constituting a combination of activities, capabilities and responsibilities[2].

The components outlined in this section serve as guiding principles for the definition of the method in this working paper.

2.2 Definition and Significance of Metadata

The concept of metadata has already been examined and outlined within the first working paper of the Competence Center (CC) for Corporate Data Quality (CDQ) (see working paper BE HSG/ CC CDQ/ 1 “Data Quality Management – Definitions and State of the Art”).

[2] For a detailed definition of a “Activity” and “Role” see the Glossary in Appendix A.

Therein we defined metadata in general as structured data that describes other data (its meaning and properties) [Burnett et al. 1999, p. 1212] 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 [English 1999, p. 482] or the general, semantically unambiguous understanding of data within an enterprise. Metadata, accordingly, facilitate the identification, detection, use and management of data as they allow an organization to better understand its data sources and definitions [cf. Marco/Smith 2006]. TOZER summarizes its function by "seeing metadata as the means by which the *structure* and *behaviour* of data is recorded, controlled, and published across an organization" [Tozer 1999, p. xix]. The most comprehensive definition is provided by MARCO, 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" [Marco 2000, p. 5]. The significance of this definition derives from its strong business orientation that we are pursuing in this working paper as well. Herein, we utilize metadata in the form of attributes (see Chapter 3.3.4) that need to be defined for specifying syntax and semantics of each integration object.

Due to their high semantic content, metadata are the fundamental components for the design of information object models as well as Business Data Dictionaries (BDD), since they serve as an original source for the definition of data elements and, hence, possess high strategic value for companies [cf. Chisholm 2006]. This strategic value primarily results from the significance the use of metadata has for evaluation and improvement of data quality by unambiguously characterizing integration objects and, consequently, increasing the comprehensibility and consistency of as well as the confidence in data [Tozer 1999, Marco 2000, Shankaranarayanan/Even 2006]. By maintaining information about the source of data or their (change) history, metadata facilitate the challenge of keeping data consistent, accurate, and complete. High-quality data, in turn, is pivotal for enabling service-oriented business applications [cf. Newman/Logan 2006], for helping to increase the validity of strategic decisions [cf. Shankaranarayanan et al. 2003], and for allowing high regulatory compliance [cf. Friedman 2006]. Moreover, metadata provide 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 enhances the decision-making quality [Marco 2000, pp. 18 ff.]. The semantic content is essentially provided by (textual) definitions of the corresponding 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 BDD [Dyché/Levy 2006, p. 164].

For the purpose of this working paper we follow the differentiation between technical and business metadata [Brunner/Dinter 2003, p. 300]. While the former defines physical data assets, the information systems infrastructure and databases and is created and used by applications managing data, we are focusing on the latter comprising conceptual descriptions and definitions of entities (i.e. a common business vocabulary) necessary for the running of the business and business users [Marco 2000, p. 49]. Business metadata includes definitions of attributes in business terms as well as business rules that apply to these attributes and data owners with the goal of helping the business user to navigate the existing information [Berson/Dubov 2007, p. 129]. We further divide business metadata into the following sub-categories for the purpose of assigning the attributes to be defined for each integration object (see Chapter 3.3.4) to a metadata cluster:

- Administration Metadata, providing general information about an entity, such as its priority, its current status, its definition process, general comments, its maintenance and instancing process etc. [Shankaranarayanan/Even 2004];
- Terminological Metadata, containing the meaning (semantics) of an entity through a unique name, its synonyms and an unambiguous definition;
- Structural Metadata, describing the structure of an entity (such as necessary attributes) as well as relationships to other integration objects [NISO 2004, p. 1];
- Governance Metadata, specifying accountabilities for each integration object;
- Compliance Metadata, designating regulations that influence or constraining the behavior and use of an integration object, but also comprising business rules, i.e. general guidelines or constraints that need to be considered for the sake of consistency when using the integration object [Tozer 1999, p. 10, von Halle et al. 2006, p. 6]. Moreover, we integrate security related metadata in this category due to their close relation to compliance issues.

- Usage Metadata, revealing how the entity is embedded, i.e. in which applications, processes and organizational units the integration object is used. Consequently, attributes comprise amongst others “Usage in Processes” and “Usage in Applications”. These metadata is particularly important for information architecture analysis and design (e.g. information flows).

2.3 Means for Defining and Organizing Metadata

Given the strategic significance of metadata for companies, a managed metadata environment becomes obivous to persistently store and technically manage data entity definitions. For realizing such a logically centralized metadata repository multiple approaches and tools exist differing by their degree of formalization and their intended use. We start with the concept of ontologies, as the notion of ontologies is the most generic and generally comprises a number of other concepts, amongst others glossaries, data dictionaries and data models, that will be broached respectively thereafter. Consequently, this generalized concept allows us to range the available approaches and facilitate the identification of their advantages and restraints.

Additionally, we would like to incorporate another concept that differs from the tools mentioned so far, namely the concept of classification systems. The reason for this inclusion results from the characteristics of coding systems, such as the United Nations Standard Products and Service Code (UNSPSC), to structure products and services in a hierarchical way and, consequently, support a common understanding of integration objects. This hierarchization allows a deduction regarding the properties of the entities that are included, e.g. what a product or service consists of and which other integration object it is part of.

2.3.1 Ontologies

The most widely accepted definition for an ontology is proposed by GRUBER who describes ontologies as a “formal, explicit specification of a shared conceptualization” [Gruber 1993, p. 199]. This definition implies three constituents: firstly, an ontology is characterized by a minimal degree of formalization, i.e., in a language whose formal properties are well understood (ideally grounded in some form of logic); secondly, it provides a common understanding of the considered subject area defining its vocabulary, i.e. the entirety of relevant concepts and their relationships [Leser/Naumann 2007, p. 273] in an abstract, simplified manner (“conceptualization”); and lastly, as the term “explicit specification” implies, the identified concepts and relationships are given concrete names and definitions [Uschold/Gruninger 2004, p. 59]. The term “shared”, finally, relates to the idea of a consensus within a certain community using the ontology, on which the conceptualization should be based.

Due to this very general definition, even glossaries are often regarded as simplistic, lightweight ontologies. However, there is a whole continuum on which different types of ontologies – including data dictionaries and data models that will be presented in the following sections – can be positioned according to their degree of formalization (see Figure 2-2). The continuum resembles from left to right the spectrum proposed by [Obrst 2003] distinguishing between weak and strong semantics. Accordingly, ontologies possess an increased semantic expressiveness the more formalized they are, making them readable and processable for machines.

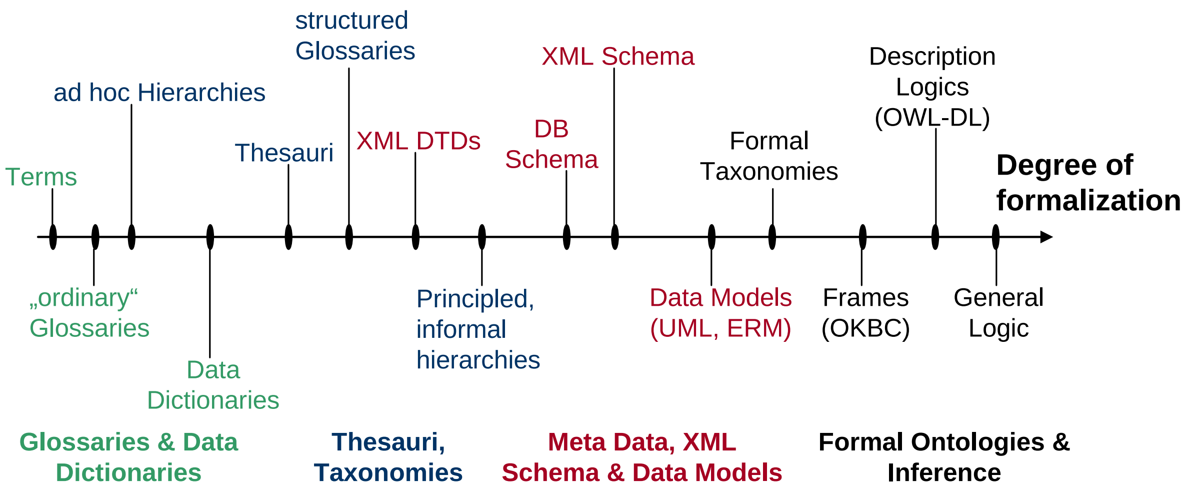

Figure 2-2: Different ontologies on a continuum ranging from lightweight to rigorously formalised (based on [Uschold/Gruninger 2004, p. 59])

Therefore, we consider ontologies as an overarching concept that comprises the three other ones.

2.3.2 Glossaries

Glossaries constitute a common approach to semantically enrich certain notions by verbally defining them. They are particularly prevalent in books for specifying pivotal concepts used in the book, but have been applied in various research fields as well. Glossaries represent a list of terms with their corresponding definitions for a particular domain of knowledge. Hence, they provide helpful explanations to a certain field of study and constitute a prerequisite for a common understanding among stakeholders. Their major deficiencies consist in a low degree of formalization (as they represent a list of textual specifications only) and the very rudimentary mapping of structural information such as relationships and interdependencies to other terms via cross references.

2.3.3 Data Models

In contrast to glossaries, data models are formal and abstract representations of data or information entities, their relationships and attributes. Their formalism is a result of the graphical notation that is used for mapping a real-world domain; in general data models are depicted in corresponding diagrams. Consequently, data models highlight structural aspects (i.e. relationships) between entities rather than their exact semantic specification and are more likely intended for implementation use. For the latter, supplementary textual descriptions are often added to the graphical data models.

Entity-Relationship-Diagrams and object models of the Unified Modelling Language (UML) are most commonly used for creating data models.

2.3.4 Data Dictionaries

Data dictionaries (also called data catalog) [cf. Elmasri/Navathe 2000] originated in the field of database management systems (DBMS) and data warehouses where descriptions of data, their relationships, access rights, constraints, data types and field lengths were stored [Devlin/Murphy 1988, p. 78]. In its most general sense, it constitutes a catalogue of metadata that contains definition, usage and representation rules of data elements. We refer to Data Dictionaries as part of the physical data modelling level including information on tables, data fields, primary and foreign keys, integrity constraints, stored procedures and triggers. The information saved in a data dictionary can then be used for analyzing data structures and the consistency of the data base [Balzert 2000, pp. 247f.]. With the objective of having centrally defined data structures, the application data of an enterprise is described.

2.3.5 Business Data Dictionaries

Within the Competence Center for CDQ we take a wider perspective on the concept of Data Dictionaries and utilize it in a more business oriented, conceptual domain as there is still little support for business metadata compared to technical metadata [Shankaranarayanan/Even 2006, p. 91]. This corresponds to the concept of so-called Information Resource Dictionary Systems whose importance for information resource management has already been emphasized by [March/Kim 1988]. It means that we integrate a larger number of attributes in order to characterize integration objects (IOs)[3] from a business perspective. Consequently, we define a Business Data Dictionary as a tool for the collection and maintenance of a set of (business and technical) metadata that contains semantically precise definitions and representations of integration objects. With this semantically unambiguous description of essential integration objects (including their relationships to other integration objects as well as technical aspects) BDDs provide a manageable possibility to generate or at least increase enterprise-wide transparency. The metadata is classified and catalogued by subject and related to other entities for ease of maintenance and retrieval. The term "Business" aims at emphasizing the conceptual scope of the tool as the represented entities should have business relevance. Summarizing its main characteristics, we detect that a BDD as a minimum requirement needs to:

- Contain a list of all essential integration objects of a company,
- Describe these integration objects with their exact semantic meaning (including their scope of application) and their relationship to each other, and
- Define the way, the integration objects are represented.

As a final remark we would like to point out that the BDDs we intend to implement within the Competence Center for CDQ are not simple glossaries or data dictionaries (as shown in Figure 2-2) but incorporate a structural and hierarchical information about the mapped integration objects. Consequently, they can be positioned at the level of "structured Glossaries" in the continuum of ontologies (see Figure 2-2) rather than on the level of traditional Data Dictionaries as described in the first part of this chapter.

[3] For a definition of the term "Integration Object" and the rationale for its introduction, please refer to Chapter 3.1.

2.3.6 Classification Systems

The term “Classification Systems” depicts coding systems that are used to classify both products and services in a standardized way. Typical examples are UNSPSC, eClass, ETIM (ElektroTechnisches InformationsModell). These classification standards defer slightly in terms of scope, geographic distribution and industries applying them. While UNSPSC constitutes a generic and globally employed standard eClass and ETIM have their widest distribution in Germany and Europe with the latter being specifically developed for the electronic industry. However, the general approach pursued by those classification systems is similar: They order products and services in a hierarchical way (top-down) with an increasing degree of granularity. Both, UNSPSC and eClass, use a four-level classification hierarchy with each level being represented by a double-digit number. Figure 2-3 illustrates the UNSPSC classification scheme with the help of an example.

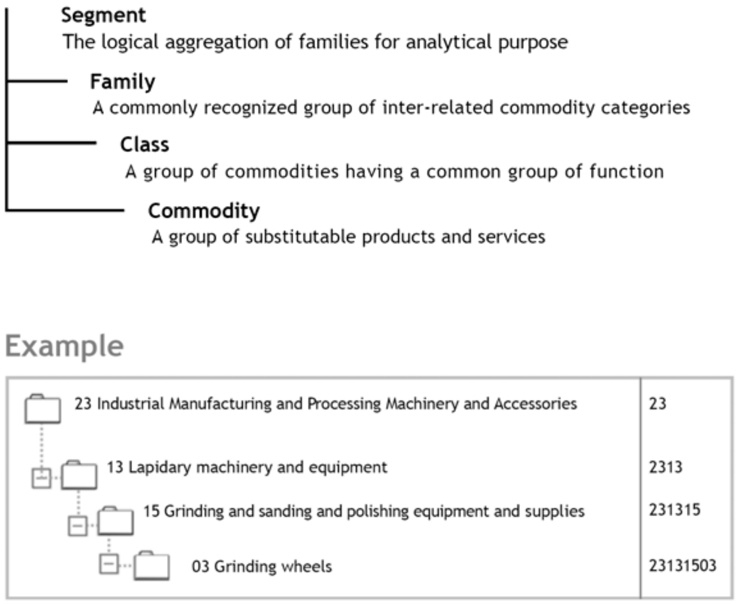

Figure 2-3: The UNSPSC coding system with a corresponding example

As we can see, such classification schemes provide a means to map structural information showing “is part of”- or “consists of”-relationships. The eClass coding system exceeds this concept by enabling users to attach characteristics to the classified entities. Consequently, it allows for a more detailed description of the mapped products (i.e. integration objects) increasing the content of semantic information [Quantz/Wichmann 2003, p. 44]. However, a comprehensive definition for each integration object is not envisaged.

2.3.7 Conclusion

As Figure 2-2 points out, the enumeration of tools and concepts for administrating metadata in this working paper is far from being complete; other means are possible but will not be broached in this working paper as they are not appropriate for our goals and for our intended target group of business users. The concepts described in the chapters before need to be evaluated by means of objective criteria in order to allow for an estimation of the appropriateness of the different tools to ensure transparency and consistency among integration objects on a company-wide level.

In several workshops with subject matter experts[4] from our project partner companies we therefore identified functional requirements for a metadata management tool from a user's perspective. This resulted in the following criteria relevant for efficient metadata storage and usage:

- the ability to capture a maximum of semantic information within the definition;
- the possibility to map relationships between integration objects, both interconnections and hierarchical relations;
- the intelligibility for non-IT users, meaning to which extent employees without profound IT knowledge can detect information contained in the tool;
- the manageability, i.e. the possibility to maintain the contained information, in terms of changes and/or extensions; and
- the integration into other systems/applications, so that the average user is able to access the required information, e.g. via a hover box, directly when he needs it (during his work).

If we compare "ordinary" glossaries, traditional data dictionaries, data models and BDDs (as we define them within our Competence Center) as well as classification systems, we obtain the following table of evaluation results.

[4] A complete list of the participants can be find in Appendix C.

Evaluation Criteria	Glossary	Data Model	Data Dictionary	BDD	Classification Systems
Semantically enriched definitions	◕	◔	◕	●	◔
Mapping of relationships	◔	●	◔	◕	◑
Intelligibility for potential users	●	◔	◑	◕	◑
Manageability, Maintainability	◔	●	◕	●	◕
User friendly integration into other systems/applications	●	◔	●	●	◕

Key: ● Very High ◕ High ◑ Medium ◔ Low ○ Very Low

Table 2-1: Evaluation of tools for administrating metadata

As we can deduce from Table 2-1 most of the concepts presented in this chapter reveal considerable shortcomings in terms of at least one of the criteria identified for evaluating their appropriateness for the goal of establishing a comprehensive and unambiguous definition of integration objects. This can be the insufficiency to include structural information (as in the case of glossaries and data dictionaries), the ability to include precise semantically enriched definitions (as e.g. data models and classification systems) or the insufficient possibility to integrate the concept for users into other applications (as in the case of data models). Finally, the low intelligibility due to rather complex and restricted representation formats represents an essential drawback of most tools [Shankaranarayanan/Even 2006, p. 91].

Within the context of our method, we therefore consider BDDs as the most appropriate means for establishing company-wide transparency of integration objects. Three factors endorse this decision:

- Business Data Dictionaries provide a balanced trade-off between a minimum of formalization and a maximal user-friendliness for employees of non-IT departments;
- They capture a large part of the necessary semantics – concise terminological definitions as well as structural information represented by data models;

- And BDDs can be easily maintained, extended, updated and interlinked with adjacent data quality management applications, such as Data Quality Scorecards and alikes.

The intention of the Business Data Dictionary is to provide anyone in a company with the definition of an integration object, no matter in which department they work or which application they are using. Therefore, the BDD must be ubiquitous – available from anywhere in the organization, at any time and to everyone. It should be easy to use, and intuitively accessible for data customers – both IT and non-IT users – in order to properly interpret and use data.

We deliberately want to constrain the scope of the BDD to a limited number of vital entities being described therein in order to preserve a reasonable size and to avoid unrealistic expectations of creating a single data model or an overarching set of universally-understood concept definitions that have already failed in the past [Rosenthal et al. 2004, p. 47].

3 METIO – A Method for Establishing Transparency on Integration Objects

The following subchapters present the pivotal artefact of this working paper: a method allowing companies to systematically establish transparency on and consistency among information entities. We refer to this method as METIO. The method will be outlined with its constituent components according to the Method Engineering metamodel (see Chapter 2.1):

- the METIO metamodel showing all relevant design objects of the method via a graphical notation (Chapter 3.1) supplemented by a glossary containing the design objects' exact specification (Appendix A),
- an enumeration of essential roles indicating necessary actors and their corresponding tasks (Chapter 3.2),
- the procedure model describing the whole process with the activities necessary to be executed (Chapter 3.3),
- the documentation model illustrating all documents created during the application of METIO (Chapter 3.4), and
- a short description of possible techniques that can be applied for the different business tasks (Chapter 3.5).

Not included in the scope of this working paper is the identification of a relevant area of application of this method, i.e. detecting the processes or organisational units for which this method is to be applied. Consequently, we assume as a basic prerequisite for the adoption of METIO that a company is aware of both a lack of transparency on its integration objects and the domain in which transparency is to be established or improved. The method is primarily intended for adoption in large-scale companies that, due to their size and international scope, dispose of a certain complexity and inconsistency in regard to their relevant information entities and possess the necessary resources to apply METIO (e.g. implement the defined roles).

3.1 Metamodel

In the research field of information systems the term “metamodel” comprises a rather heterogeneous spectrum of models [Strahringer 1996, p. 12]. We therefore need to differentiate between “general” metamodels, which define other models or modelling concepts [Frank 1994, p. 172] and metamodels as an integral part of a method. Hence, for the working paper at hand two definitions appear most adequate: Firstly, FERSTL and SINZ describe metamodels as a framework that specifies the types of modelling entities (object types) and relationships between these entities along with their semantics and usage rules [Ferstl/Sinz 2001, p. 88]. Secondly, and more precisely, RUMBAUGH defines metamodels for a method as a description of the concepts in the method and their relationships to each other [Rumbaugh 1995, p. 12]. These two definitions correspond to our understanding outlined in Chapter 2.1, suggesting that a metamodel should map relevant entities of the method, i.e. design results generated in corresponding design activities, as well as the relationships between them as a conceptual data model [Gutzwiller 1994, p. 14]. Figure 3-1 depicts the metamodel for the METIO method. The respective entities of the metamodel, as well as their relationships, are described in more detail in Appendix A.

The METIO metamodel can be subdivided into three parts or domains that focus on different subject areas. Firstly, the metadata definition part (highlighted by the simple framed rectangle in Figure 3-1) contains all entities relevant for the identification and definition of integration objects. The metadata of these general integration objects as well as those of the relationships between them are stored (in our case) in the Business Data Dictionary. The integration object instances ought to be created in accordance to the IO definition.

Secondly, we map the operational and organizational structure in which the method is embedded, particularly the (meta-)data definition and usage processes as well as their manifestation in organizational units and roles (see double framed rectangle in Figure 3-1). More specifically, the integration object metadata stored in the BDD constitute the output of the corresponding business tasks of the metadata definition process. This output, in turn, serves as input for metadata usage processes as well as data usage and maintenance processes when concrete data instances are used in the operational work. From an organizational point of view we have defined roles (that are specified in Chapter 3.2) who provide a consistent and transparent definition and maintenance process of relevant integration objects. The characterization of these roles as well as their corresponding business tasks give us a link to the Data Governance topic, broached in more detail in the working paper BE HSG/ CC CDQ/ 3.

And thirdly, the Data Quality part (highlighted with the dashed framed rectangle in Figure 3-1) allows the integration of the definitions for quantification and measurement that can be used for evaluating the excellence and consistency of object definitions by means of a Data Quality Scorecard. As already mentioned at the beginning of this paper (see Chapter 2.2), metadata, and hence a BDD that contains such data, plays a significant role in evaluating and improving data quality as it provides specifications for the proper definition and use of a company's essential information entities. In this fragment of the metamodel the compliance of instantiated business and data objects with their corresponding IO definition given in the BDD are determined (based on certain metrics such as the completeness of specified attributes) and compared to predefined target values. With the help of such metrics the quality of the used data can be more easily measured and monitored over a period of time. This part of the metamodel is intended to establish a direct link to the subject of Data Quality Scorecards as a tool for data quality measurement and monitoring. However, this topic will not be in the focus of this working paper, but is subject to future research within the CC CDQ.

An important relationship that explicitly needs to be highlighted at this point of the paper is the relationship between the two entities "Data Quality Variance" and "Metadata Definition and Maintenance Process". It does not only constitute a pivotal link between the data quality part of the metamodel and the organizational processes and structure, but also illustrates the fact that poorly defined integration object instances need to trigger a redefinition of the integration objects concerned. Similar triggers that initiate a metadata refinement out of operational data management processes are further elaborated in Chapter 3.3.3.

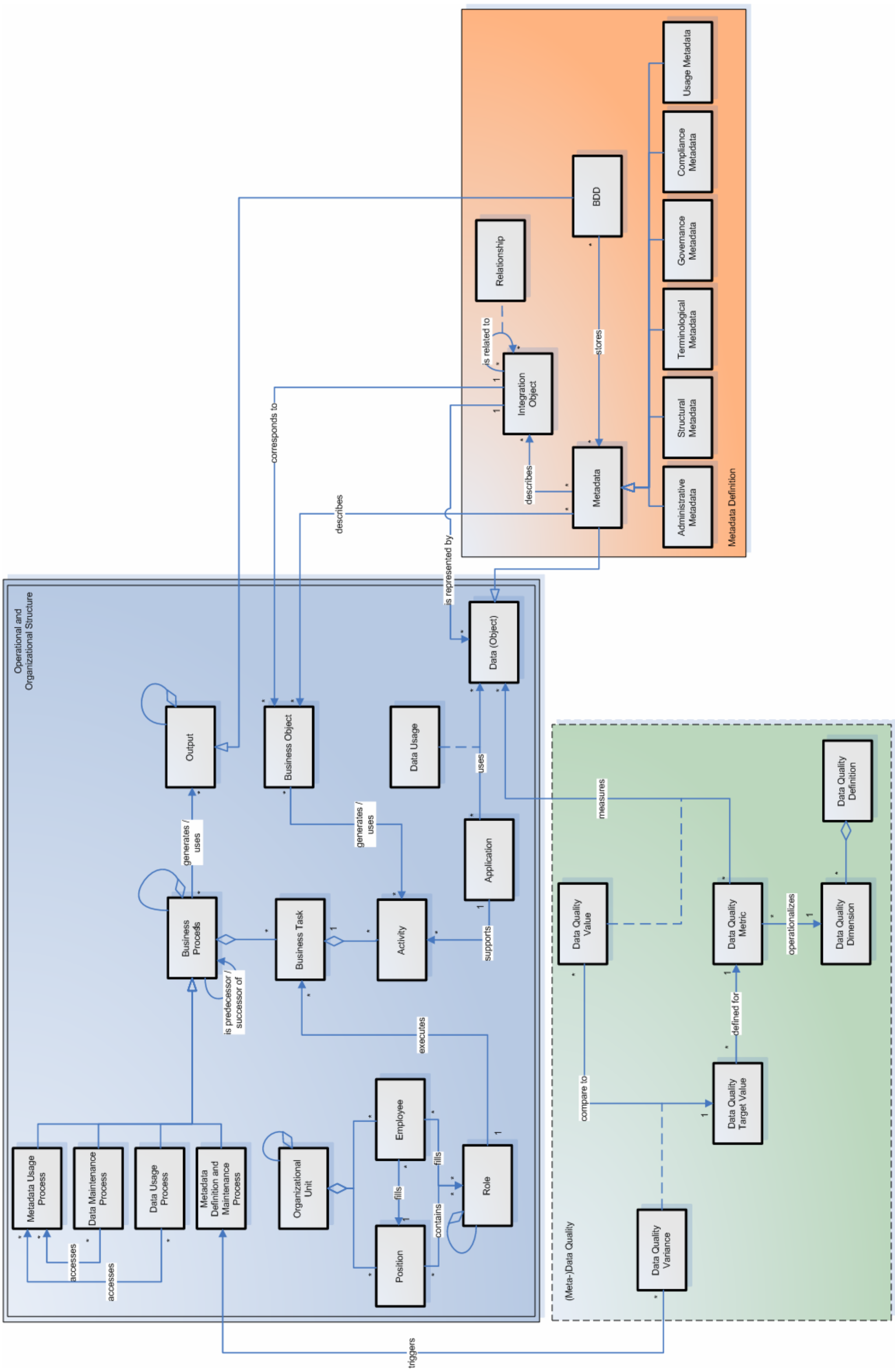

Figure 3-1: Metamodel of the METIO method

Referring to integration objects, we would like to clearly demarcate the concept from related terms, such as business objects and data objects or data elements[5]. The differentiation will be of great importance in the following sections. Within this working 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. According to the OMG, they represent a thing active in the business domain, such as a person, a business process or a concept and are implemented by one or more objects in the information systems [OMG 1996, p. 19]. These business objects are relevant to business experts and generally described (if at all) in a simple textual form (such as a glossary) or a list of their constituent attributes – similar to the business object description proposed by [Scheer 2001, p. 193]. The definition contains a coarse-granular description of characteristics relevant from a business perspective. Data objects on the other hand, are coherent, self-contained pieces of information suited for automated processing [Lankhorst 2005, p. 99], i.e. 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 an 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 containing integration objects that constitute business relevant entities on a logical level. This integration object layer represents a layer of reference on which process-specific business objects and application-specific data objects are mapped. Integration objects are described with their semantics as well as with their structure (consisting of relationships to other integration objects) and go beyond purely business oriented definitions[6]. They are modelled and described with their entire set of attributes and consequently represent an integrated view of both business and data object characteristics. By contrast, data objects are application-specific storing a subset of characteristics of the corresponding integration object. Figure 3-2 illustrates the described distinction on three layers.

The rationale for introducing the concept of integration objects in addition to business and data objects results from the need of a consolidation between business and IT. Integration

[5] The terms data object and data element are used synonymously in the rest of this working paper.

[6] For a further definition we refer to the Glossary of the METIO metamodel in Appendix A.

objects represent the technical representation of the real-world business objects, including technical attributes that are valid across applications. These consolidated entities in terms of a “least common denominators” of the different attributes become particularly relevant when modelling information flows between applications within the scope of architecture analysis and development. The concept of integration objects also accommodates the fact, that unlike data (as a simple structured set of values stored in a system) information is characterized by being placed in a particular context [Tozer 1999, p. 31]. Therefore, integration objects need to be described with the processes, applications and organizational units using them (see Figure 3-6).

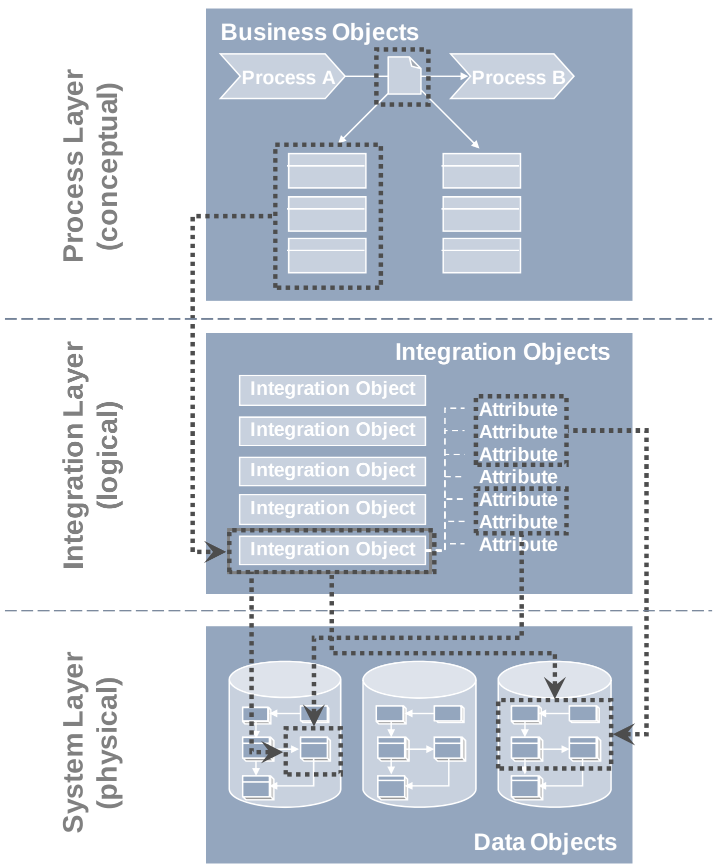

Figure 3-2: Three-layer differentiation between business, information and data objects

The tripartite differentiation corresponds to general approaches pursued e.g. by the Object Management Group in the Model Driven Architecture (MDA). Within this framework the OMG distinguishes between three different viewpoints and, hence, models. In conformity with our three-level structure, the Computation Independent View focuses on the environment of a system (processes in our case) independent from the concrete implementation with the domain practitioner as the central target group [OMG 2003, p. 15].

This corresponds to our business objects. On the level below, the Platform Independent Model, the system is still considered on a conceptual level unaffected of the underlying platform, but already specifying concrete operations of the system. And finally, the Platform Specific View characterizes in detail how a system uses a particular type of platform [OMG 2003, p. 16].

3.2 Roles

The roles of the METIO method are specified in accordance to the roles defined in the Data Governance working paper ("Design Criteria for Data Governance", Working Paper BE HSG/ CC CDQ/ 3). The responsibilities and tasks have been adopted as far as possible. However, aspects were added that are specific to the definition and harmonisation of integration objects [cf. Marco/Smith 2006].

This chapter consists of an enumeration of the roles that take part in the whole process of establishing and keeping transparency on integration objects as outlined in the procedure model in the following chapter. For each of the roles we add a short specification of the responsibilities and tasks these roles need to fulfil within the METIO procedure model (see next chapter). In order to avoid repetition we will only concentrate on METIO specific tasks and refer to the above mentioned working paper for more general activities.

- *Business Data Steward.* A Business Data Steward works directly with representatives from business. He documents business requirements and assess the impact of new business requirements on data and vice versa. His tasks may involve creating business rules for data, developing data models and data vocabularies, implementing data management best practices, and maintaining and publishing data quality metrics. Business Data Stewards know how business terminology is defined in their area and how business processes use data. Therefore, they are responsible for identifying business objects and unambiguously describing business oriented attributes of relevant integration objects. The Business Data Stewards that are assigned by the Chief Data Steward to this task (see Chapter 3.3.2) become automatically primary owners of the respective entities and, hence, responsible for the maintenance of the metadata and the processing of change requests concerning their integration objects.

- *Technical Data Steward.* The counterparts of Business Data Stewards are Technical Data Stewards, who focus on data's representation in IT systems. For their area of responsibility, they provide standardised data element definitions and formats and focus on technical metadata. A metadata repository may record those standards. In addition, technical stewards profile source system details and data flows between systems.

- *Chief Data Steward.* The Chief Data Steward often chairs the Data Governance Board. His main task is to put the board's decisions into practice. The Chief Data Steward is an expert in data quality issues across the enterprise. Within METIO he plays a vital role as he does not only ensure that all relevant integration objects are included for definition but also assigns those entities to the corresponding Business and Technical Data Stewards. He consequently can be considered as the pivotal role and responsible for the metadata management.

- *Data Governance Board.* The Data Governance Board defines the data governance program for the whole enterprise. It may guide related initiatives like data architecture, data integration, data warehousing, metadata management, and master data management. As final and highest decision making authority for the unambiguous definitions of integration objects he is accountable for a consensus-driven establishment of a uniform understanding.

- *Process Owner.* The Process Owner is the responsible for the process in which a certain business object is created, maintained, used or exchanged. Due to his knowledge of the specific process, and hence the context in which the business object is embedded, he is consulted in the process of identifying relevant business objects.

- *Process Engineer.* Just like the Process Owner, the Process Engineer is consulted to bring in his knowledge about the process and the included business objects to assist the identification and description of relevant business objects.

- *System Owner.* On the system side the System or Application Owner is the responsible for a certain IT system and, consequently, for the data objects created, maintained, used, saved and deleted by an application or exchanged with others. This specific knowledge is used to identify and describe relevant data objects.

- *Data Engineer.* A Data Engineer is concerned with the creation and modification of data sources as well as the definition and execution of transformations within a company. During the process of identifying relevant data objects he plays a consultant role. The notion is intended as a generic role subsuming related roles such as a Data Owner, Data Analyst or a Data Modeller.

- *Security Agent.* The Data Security Agent is responsible for classifying security-sensible data and, therefore, defines whether such data is confidential or allowed for internal use only.

- *Legal Department.* The Legal Department adds compliance-relevant metadata to an integration object if necessary.

3.3 Procedure Model

The procedure model of METIO was partly derived from established approaches in information systems analysis and engineering [Ferstl/Sinz 1995, Balzert 2000] serving as rough guiding principles, enhanced with experience gathered during adoption of METIO in some of the companies from our Competence Center (see Chapter 4). Moreover, we did not only discuss and refine the approach in iterative discussion with our project partners[7], but investigated and analyzed approaches from companies that either conducted projects in their own company (such as Audi) or in other companies (such as Alfabet, a software provider specializing in enterprise architecture analysis and design) pursuing similar goals.

Figure 3-3 illustrates the overall procedure model for establishing company-wide transparency on essential integration objects. The method constitutes an iterative approach with regard to the scope of application (from a single process to company wide application) which accommodates the fact that it is illusory to create an overarching set of universally-understood concept definitions ad hoc and therefore advisable to "standardize in small granules" [Rosenthal et al. 2004, p. 47].

[7] For further information on the participating experts we, again, refer to Appendix C.

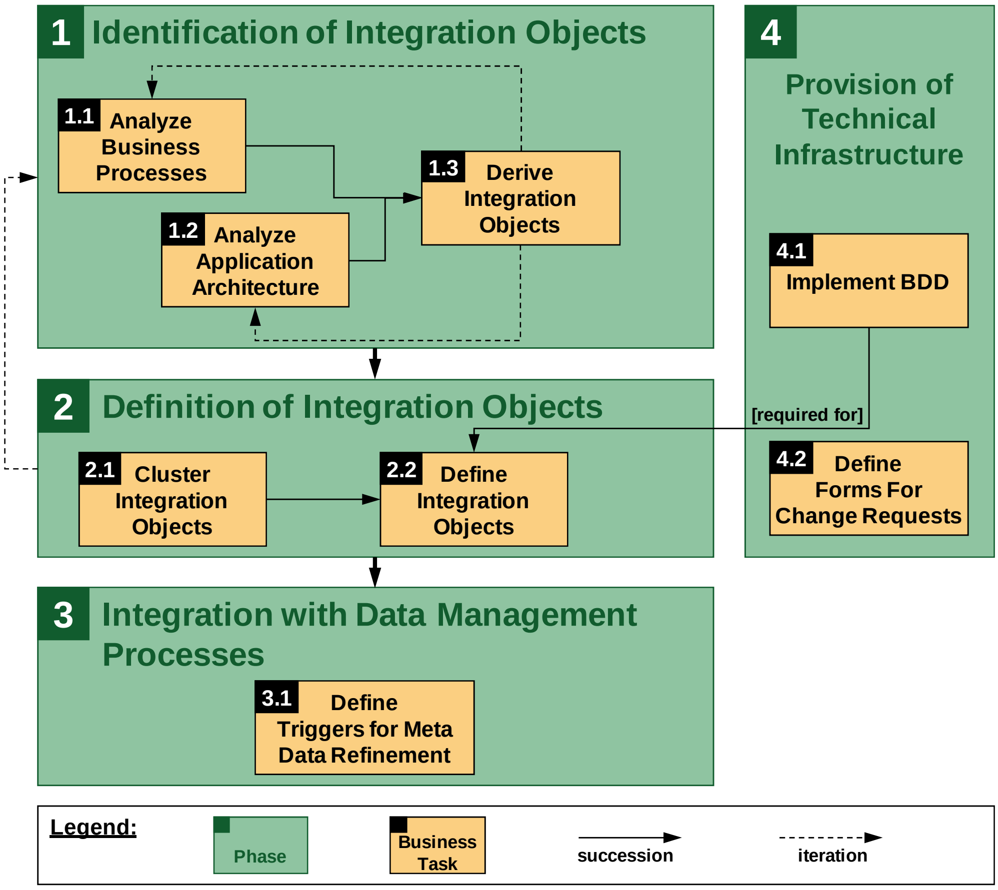

Figure 3-3: Overall Procedure Model of the METIO Method

The procedure model will be described in detail in this chapter. For the sake of a uniform description we use the following structured template that is derived from the Method Engineering metamodel outlined in section 2.1 in order to specify each of the constituent business tasks.

<table>
<tr><td colspan="3">Business Task
In this box the business task is given an unambiguous and meaningful name that allows to distinguish it from other business tasks.</td></tr>
<tr><td>Output
The most significant output(s) of the business task are specified.</td><td>Role(s)
The roles necessary to conduct the business task are stated in this cell.</td><td>Technique(s)
Possible techniques that help achieving the intended output of the business task are presented.</td></tr>
<tr><td colspan="3">Description
This part of the template contains a textual description of the business task, including their main objective.</td></tr>
<tr><td colspan="3">Activities
Based on the description represented above, the activities necessary to complete the business task are enumerated and characterized in this section.</td></tr>
</table>

Table 3-1: Template for the description of business tasks within the procedure model

In subchapters 3.3.1 to 3.3.4 the different phases of the METIO procedure model (see Figure 3-3) will be characterized with their corresponding business tasks according to the description template. As integral parts of a method, outputs or design results (such as resulting documents), roles and techniques are further detailed in the following chapters. The roles are additionally detailed by participatory responsibility types with the help of the RACI model that was originally developed for the IT Governance reference framework COBIT [IT Governance Institute 2005], where:

- R stands for Responsible for the people that actually do the work and take decisions,
- A signifies Accountable, i.e. for the resource authorizing decisions and, hence, being ultimately accountable for the completion of tasks,
- C denotes Consulted for roles that are asked to provide input and support decisions,
- I means Informed, i.e. people that are just informed of the decisions.

For each business task exactly one responsible role needs to be assigned; the lack of a responsible or a assignment to several roles has to be excluded in order to avoid ambiguities in terms of business task performance.

3.3.1 Identification of Integration objects

To a large extent, the first part of the METIO procedure model is based on established approaches from systems / requirements analysis [Krallmann 1996, Balzert 2000, Sommerville 2007]. The analysis process described therein was adapted to the specific requirements of our case and supplemented with corresponding roles and techniques.

Business Task Analyze Business Processes		
Output	**Role(s)**	**Technique(s)**
List of essential business objects (including prioritization)	o Business Data Steward (R/A) o Process Owner (C) o Process Engineer(s) (C)	o *Business Process Analysis*[8] o *Business Object Analysis* o Identification of business objects by *content analysis* of business process documentation o Identification of business objects through *group and individual interviews* with

[8] Techniques written in italic are further specified in Chapter 3.5.

		process owners
Description This first business task intends to identify relevant entities from a process perspective. Therefore, the already existing process documentation needs to be worked through. 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 are necessary to obtain the missing information and essential business objects.		
Activities o Analyze process documentation, particularly process models if available o Analyze exchanged business documents and included entities o Conduct interviews with process owners to obtain additional information about the business processes o Prioritize Business Objects		

Table 3-2: Description of the business task "Analyze Business Processes"

The execution of the last but one activity in the table above becomes necessary if the documentation of the business processes – and particularly the integration objects exchanged within and/or between them – is insufficient or deficient as it is the case in many companies today.

Business Task Analyze Application Architecture		
Output List / model of data objects that are used within and exchanged between applications (including prioritization)	**Role(s)** o Technical Data Steward (R/A) o System Owner (C) o Data Engineer(s) (C)	**Technique(s)** o *Content analysis* of existing documentation (data models, applications) o *Information/Data Flow Analysis* o Additional *group and individual interviews* with system owners and data engineers
Description This business task serves as complement to the identification of the business objects in the business task before. As 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.		
Activities o Model system-specific physical data models o Consolidate physical data models and model consolidated, company-wide data architecture o Prioritize Data Objects		

Table 3-3: Description of the business task "Analyze Application Architecture"

These first two business tasks can be partly carried out simultaneously. The rationale behind these business tasks is to have a combined top down and bottom up approach that allows us to integrate essential entities both from a process and a system perspective. The most substantial entities derived from these analysis tasks are either identifying data objects without an equivalent business object on a process level, or, the other way around, a 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 recommendable.

In terms of a workload reduction, however, it can be reasonable in some cases to constraint the application or system analysis to the data objects that are directly representing the identified relevant business objects from the first business task. This allows a focus on those entities that are actually relevant from a business perspective. Such an approach was pursued at our project partner Daimler AG. The drawback of such a "restricted" analysis is that potentially redundant data objects cannot be identified properly.

The prioritization of both the business and the data objects in the last step of these two business tasks constitutes an important prerequisite for the following phase in which the consolidated integration objects are defined. The order of the integration objects to be defined should be oriented on the prioritization in order to avoid the description and modelling of every single entity but to really restrain the BDD to the most relevant ones. The rating is to be carried out by the responsible roles for the first business tasks. For the purpose of the prioritization an already established categorization standard for business data, such as the Storage Networking Industry Association (SNIA), should be applied. The SNIA defines criteria with regard to the criticality of data and distinguishes between five different categories [Thome/Sollbach 2007]:

- *Class 1: Not Important to Operations*. Data that can be restored with little effort / at low cost or duplicates of existing data.
- *Class 2: Important for Productivity*. Data used for daily business that can be restored with little effort / at low cost or substituted through alternative data.
- *Class 3: Business Important Data*. Data used for daily business and representing part of the company's business know-how.
- *Class 4: Business Vital Data*. Data that is relevant for the planning and management of a company and whose loss can lead to severe corporate damage.

- *Class 5: Mission Critical Data*. Data needed for the essential business processes and whose loss may lead to a severe operational damage, meaning that the company's goods and services can no longer be generated. Data that needs to be stored due to regulatory requirements.

The identified business and data objects should be assigned to one of these categories depending on their criticality for the company. At the beginning each company applying METIO should specify the integration objects of which categories it actually needs to define in order to describe only the relevant integration objects (e.g. Classes 3 to 5) and avoid describing all entities (such as Class 1 entities) resulting in an Enterprise Data Model.

<table>
<tr><td colspan="3">The identified business and data objects should be assigned to one of these categories depending on their criticality for the company. At the beginning each company applying METIO should specify the integration objects of which categories it actually needs to define in order to describe only the relevant integration objects (e.g. Classes 3 to 5) and avoid describing all entities (such as Class 1 entities) resulting in an Enterprise Data Model.Business Task
Derive Integration Objects</td></tr>
<tr><td>Output
Consolidated and non-redundant set of integration objects that need to be defined (including prioritization)</td><td>Role(s)
o Chief Data Steward (R)
o Business Data Stewards (C)
o Technical Data Stewards (C)</td><td>Technique(s)
o Mapping of business and data objects</td></tr>
<tr><td colspan="3">Description
At this point of the procedure model the previously elaborated results are used as input in order to derive a consolidated and non-redundant set of integration objects. Therefore, possible misfits between the process perspective (business objects) and the application perspective (data objects) need to be resolved.
After approval by the Data Governance Board, this final set of relevant integration objects is cleared for the subsequent metadata definition process.</td></tr>
<tr><td colspan="3">Activities
o Consolidate the entities from both views on a conceptual level (including their prioritization) → Identify relevant integration objects
o Approve list of identified integration objects</td></tr>
</table>

Table 3-4: Description of the business task "Derive Integration Objects"

This last step of the first phase cannot be detached from the previous business tasks. Firstly, they provide the required input for consolidation, and secondly, a return to the previous activities might be necessary in case further entities are detected.

3.3.2 Definition of Integration Objects

Business Task Cluster Integration Objects		
Output Document for each Integration object to be defined with its corresponding categories (see Figure 3-7)	**Role(s)** o Chief Data Steward (R/A) o Business Data Steward (C) o Technical Data Steward (C)	**Technique(s)** ---
Description In order to constraint the effort for the definition of the integration objects, the Chief Data Steward needs to cluster the identified entities based on their similarities. For each of these integration object clusters a separate definition process variant is defined (see following business task) depending on the roles that are necessary to be included in the definition process.		
Activities o Identify different categories of integration objects (by clustering integration objects with similar characteristics) o Assign integration object categories to metadata definition process variants and defined roles		

Table 3-5: Description of the business task "Cluster Integration Objects"

The preferable output of this first business task of the integration object definition process would be a document revealing the integration objects identified during the first phase, the different categories these integration objects belong to, the corresponding definition process variants for the integration object categories (see Figure 3-4) and the assignment of the roles responsible for effecting the definition (filled in by the Chief Data Steward at the beginning of the definition process). Such a document (see Figure 3-7) helps the Chief Data Steward to survey and structure the totality of relevant integration objects that need to be specified.

Business Task Define Integration Objects		
Output Completed and approved BDD with definitions for all relevant integration objects	**Role(s)** o Chief Data Steward (R) o Business Data Steward (C) o Technical Data Steward (C) o Data Governance Board (A) o Data Security Agent (optional) o Legal Department (optional)	**Technique(s)** o Conduct an approval *workshop* (at the end of the definition process)
Description This central business task for establishing transparency on integration objects comprises the actual definition of the essential entities. Therefor, possible definition process variants depending on the integration objects cluster are possible (see Figure 3-4). The predefined process variants as well as the executing roles are assigned in a first step to each of the integration objects. An appropriate document (see Figure 3-7) helps the Chief Data Steward for this assignment. Each of the integration objects can be specified according to the process variant it is assigned to. When the definition of an integration object is finalized, an intensive review of the definition by the Chief Data Steward is necessary in order to ensure completeness of the specification as well as consistency with other definitions. Finally, the Data Governance Board needs to approve and clear the definition based on the		

CDS's evaluation and an additional joint review.
Activities o Appoint responsible Business and Technical Data Steward(s) o Provide Usage and Terminological Metadata as well as attributes for integration objects[9] o Define relationships to and dependencies from other integration objects o Specify Configuration Requirements o Define security level (optional) o Add compliance-relevant data (optional) o Define Maintenance and Instancing Procedures for integration objects o Provide Administrative Metadata for integration objects. o Review definition of integration objects (particularly with regard to completeness) o Approve and clear integration object definition

Table 3-6: Description of the business task "Define Integration Objects"

The whole sequence of activities with their corresponding responsibilities (swimlanes) is illustrated in Figure 3-4. In a first step, the Chief Data Steward needs to appoint the Business and Technical Data Stewards responsible for the definition of the integration object metadata as well as those that need to be consulted in case of consolidation needs concerning integration objects that are used across multiple organizational units and applications. It is important to point out that the Business Data Steward assigned to the definition of a certain integration object becomes automatically responsible for the maintenance of the definition in the future and needs to look after the processing of possible change requests (see Chapter 3.3.3). Potential candidates for the role of a steward should preferably fulfil a number of desired characteristics, such as regularly interacting with or even managing the entity, having the authority to make decisions with regard to the object, having a profound understanding of the business or technical domain in which the object is used. For the assignment of responsibilities, the Chief Data Steward can use the corresponding document "IO classification and assignment" (see Figure 3-7) where he can fill in the names of the Business and Technical Data Stewards respectively who need to carry out parts of the definition or at least are to be consulted in order to find a consent decision. For this purpose the Chief Data Steward needs to possess sufficient knowledge about where integration objects are needed and who might be the right contributors for the definitions.

[9] These attributes refer to the BDD data model (see Figure 3-6) and are specified in Appendix B.

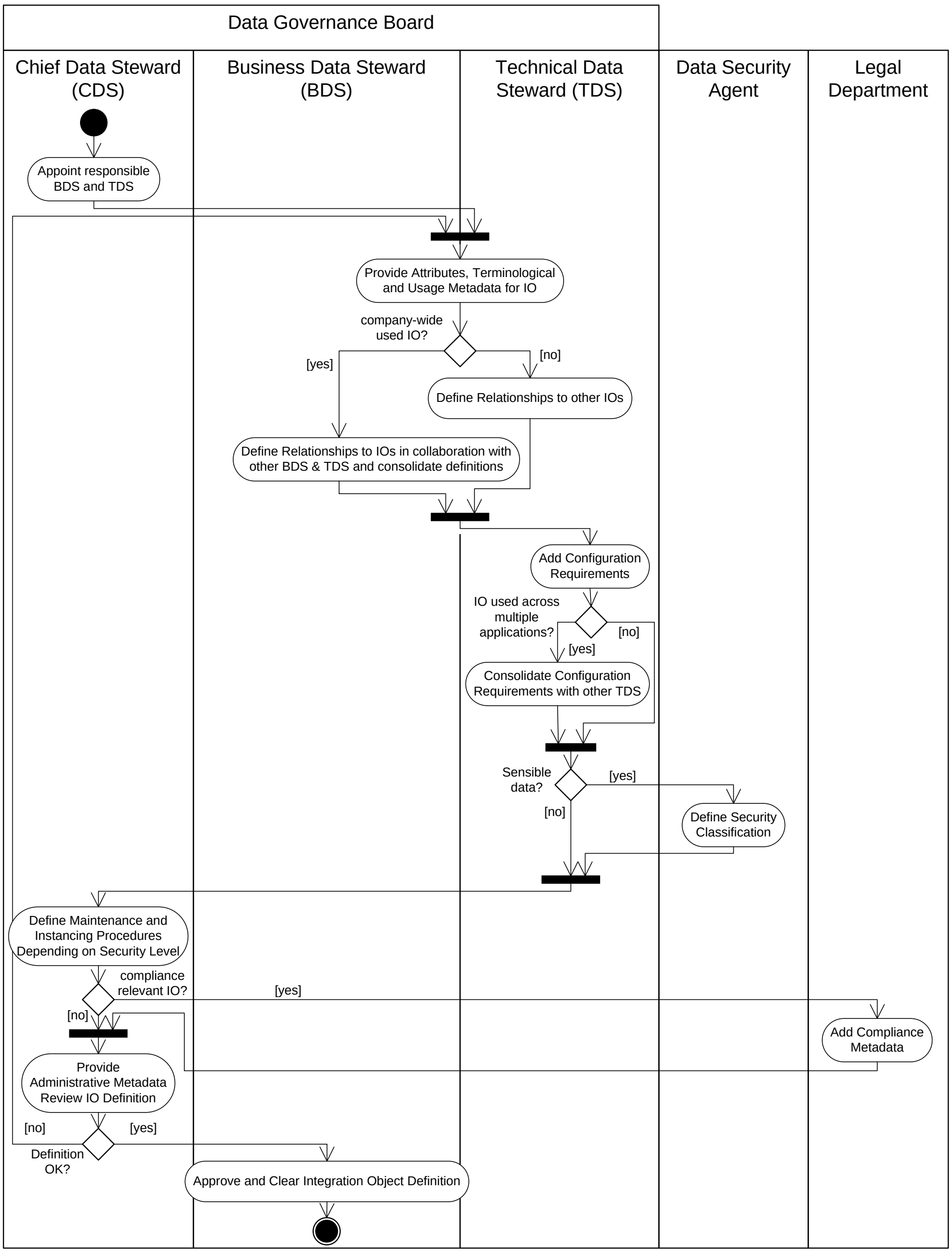

Figure 3-4: Activity diagram for defining integration objects[10]

The final two activities of the business task "Define Integration Objects" are intended to emphasize the need for governance that emerges when integration objects are to be defined on an organization-wide level. Such control is needed to consolidate terms on a cross-

[10] The names of the attributes that the different roles need to define correspond to the classes of the BDD prototype metamodel (see Figure 3-6). The location of the diamonds representing decisions within the definition process does not imply that the corresponding role of the swimlane is responsible for the decision. Instead these decisions are taken beforehand by the Chief Data Steward when the integration objects are classified.

functional level with other groups in the organization who may use these terms differently [cf. O'Neil 2005b].

It is important for the visibility of the Chief Data Steward to actively communicate the finalization of the definition process for integration objects to the corresponding process and data owners. This communication should include the rationale in case of integration objects types that were not used uniformly within a company in order to enhance the consensus among the different stakeholders and increase a uniform use in the future.

3.3.3 Integration with Data Management Processes

This process step is intended to guarantee the embedding of the metadata definition and maintenance process with the operational (meta-)data usage processes. Transparency on business and data objects within a company cannot be brought about by simply defining all relevant entities in an unambiguous way one time. In fact, it is equally important to ensure that the transparency and consistency established is surveyed continuously in order to maintain a good quality of the integration objects defined. 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 the business environment changes, this may lead to definitions which may 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 integration objects or add new ones in the most flexible way possible [cf. O'Neil 2005b]. This includes the possibility to make sure that conflicts are resolved rapidly and definitions are kept.

For this reason we consider the process of establishing and maintaining maximum transparency with the help of metadata as a nested and iterative process within the regular processes on a data level as illustrated in Figure 3-5.

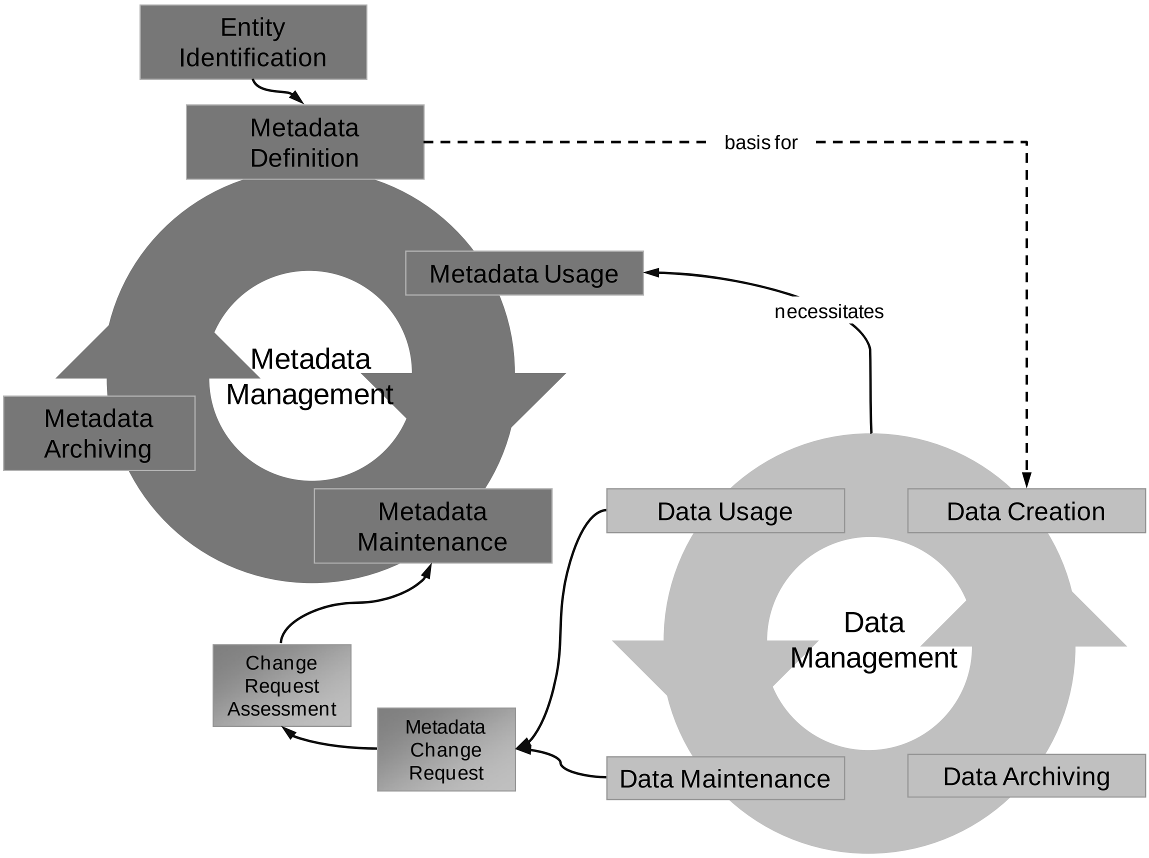

Figure 3-5: Metadata and data management in a conjoint lifecycle

This means that the triggers for transition to and from the metadata definition process (from the dark grey cycle to the nested light grey one) 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 integration object definitions, or integrate new ones are checked regularly. In accordance with the general data governance process, it appears reasonable to have these decisions within the scope of the Data Governance Board meetings taking place periodically[11].

Business Task Define / Adapt triggers for metadata refinement		
Output List of triggers leading to a necessary refinement of existing integration object metadata (see document template "Triggers for Metadata Refinement")	**Role(s)** o Data Governance Board (A) o Process Owners/Engineers (C)	**Technique(s)** ---

[11] Here again, we refer the interested reader to the working paper broaching the topic of Data Governance (BE HSG/ CC CDQ/ 3).

Description With this business task the Data Governance Board needs to define the triggers that activate a new run-through of the METIO procedure in order to adapt integration object metadata or add new ones. Possible triggers are insufficient or incorrect, i.e. ambiguous, definitions, new entities that are not yet defined in the BDD and the need to included, or the status change (e.g. retirement) of an integration object that become obsolete or is cleared by the reviewers. The triggers need to be integrated as checkpoints into existing IT and data management processes and therefore jointly agreed with responsible process owners and engineers.
Activities o Define possible triggers with threshold values o Specify ways for measurement and responsibilities

Table 3-7: Description of the business task „Define/Adapt triggers for metadata refinement"

The activities that need to be executed in this business task lead to the completion of the document template "Triggers for metadata refinement" (see Chapter 3.4). This process is highly company-specific, however, Figure 3-8 includes some possible examples for triggers as well as corresponding thresholds and ways for measurement. The completed document is to be revised regularly on the Data Governance Board meetings in order to guarantee adequately defined triggers and thresholds.

In addition, each Business Data Stewards should periodically – e.g. on a half-year basis – review the integration object metadata he is responsible for (irrespective of existing change requests). For this task he can consult the responsible Technical Data Steward and other concerned Business Data Stewards.

3.3.4 Provision of Technical Infrastructure

Business Task Implement Technical Infrastructure for BDD Tool		
Output Implemented BDD Tool infrastructure allowing to map and manage integration objects and to characterize in all of the necessary attributes	**Role(s)** o Chief Data Steward (A) o IT department (R)	**Technique(s)** o *Requirements Analysis* o *Software Testing* o *Vendor Evaluation and Tool Selection*
Description As different user groups will define, maintain and use the defined integration objects, it is recommendable to generate different views, i.e. assign different access rights for certain input fields and, hence, attributes.		
Activities o Implement BDD Tool according to BDD class diagram (shown in Figure 3-6) o Adapt BDD Tool to company-specific requirements o Define different views on the BDD Tool for different user groups o Assign roles / access rights to users o Train users on BDD Tool		

Table 3-8: Description of the business task "Define Integration objects"

Within the Competence Center for CDQ we implemented a first, web-based BDD prototype[12]. The underlying BDD metamodel, representing all relevant attributes that can be specified for an integration object definition, is shown in Figure 3-6 (mapped as a UML class diagram). The attributes are clustered (in classes) to a large extent according to the roles that will be responsible for defining them. The different metadata classes that have already been modelled in the METIO metamodel (see Figure 3-1) are illustrated by the distinct colours, distinguishing between Administrative, Terminological, Structural, Usage, Governance and Compliance Metadata as introduced in Chapter 2.2.

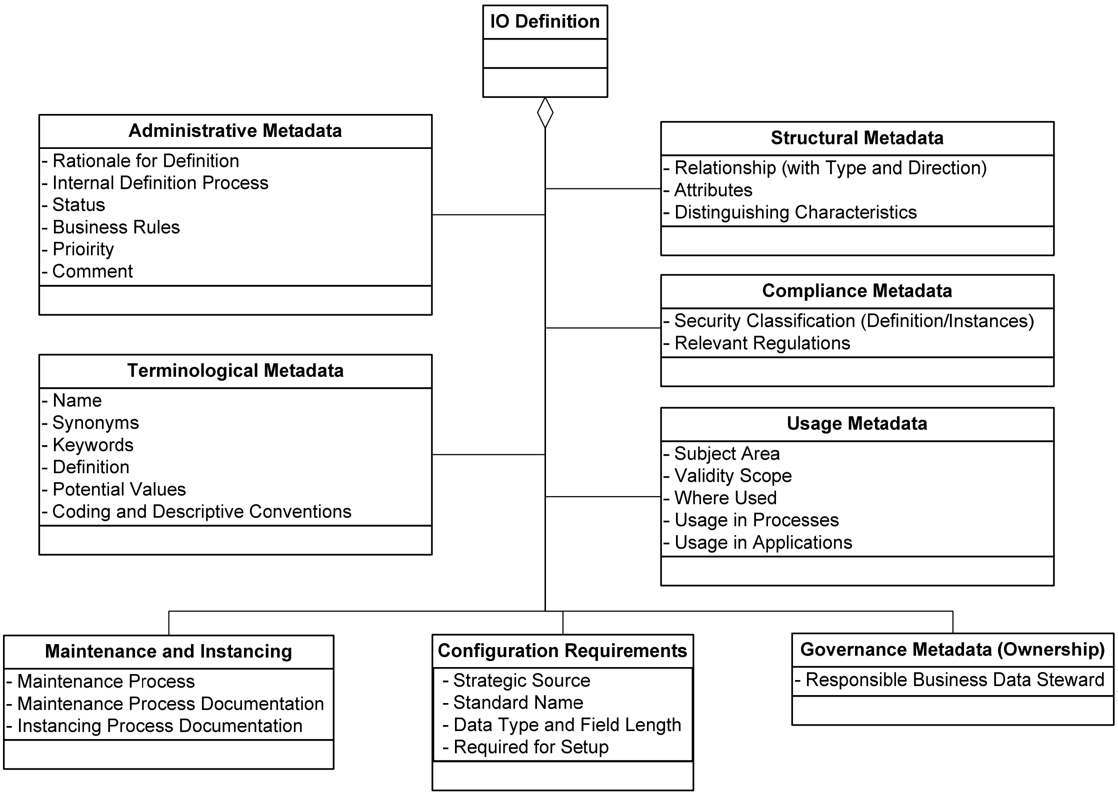

Figure 3-6: Metamodel of the BDD prototype implemented in the CC CDQ

For the IO definitions in our BDD 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 integration object. Therefore, we conducted an

[12] The Business Data Dictionary Tool is accessible for all project partners under the following link: https://secureapps-test.eta.ch/bdd/default.aspx.

in-depth literature analysis including metadata standards in adjacent research fields such as computer as well as library and information science where metadata plays 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.

Element name	Description	BDD 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 comments	Justification; why is this metadata element needed? How is it used? Other comments or instructions.	o Rationale 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	Where Used
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 3-9: Attributes to describe metadata according to [Päivärinta et al. 2002, p. 1157] and their equivalent BDD attributes

PÄIVÄRINTA ET AL. 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 [cf. Päivärinta et al. 2002]. 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 (Dublin Metadata Core Element Set) [Dublin

Core Metadata Initiative 2008], the Australian Government Locator Service [Australian Government Locator Service 2002] and the ISO/IEC 11179-3 and 4 (specification and standardization of data elements and formulation of data definitions) [ISO 1995, ISO 2003]. The identified metadata elements are summarized in alphabetical order in Table 3-9. Besides, we included a short description and their equivalents from the BDD class diagram (see Figure 3-6).

As Table 3-9 indicates, we used the majority of the identified attributes for our BDD, 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 case.

A similar synthesis was conducted by O'Neil for the components necessary to provide a sound definition within a glossary [O'Neil 2005a, p. 2]. From this list of attributes we identified a number of further components that we could use for the BDD 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 "Related Term" that – in our case – incorporates the relationship to other integration 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 his naming (to "Strategic Source") and meaning, signifying the original source of the integration object rather than the source where the definition came from;
- "Approval Information" as an attribute for tracking the governance trail was renamed to "Internal Definition Process" but kept with the same semantic meaning (comprising the approval process for the definition); and
- "Distinguishing Characteristics" and "Synonyms" were directly transferred to our own BDD model with the definitions as stated in [O'Neil 2005a, p. 2].

The element "Replaced by" was conceptualized broader and realized in a slightly different way. It constitutes a possible value within the BDD attribute "Status" (with Draft, Final and

Retired being other possible status values). In case of a “Replaced” status of an integration object a relationship “replaces/replaced by” has to be assigned to the attribute “Related Terms” (see Appendix B) 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 BDD and had to be either adapted to our specific needs (as outlined above) or completed by further attributes. For this purpose we integrated the information gathered from interviews and discussions with domain experts from our Competence Center 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, “Required for Setup” 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 BURNETT ET AL. [Burnett et al. 1999]. 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 BDD 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 integration 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 similarities as regards content and role assignment for the definition process. Moreover, the justification for the categorization results from the classification of metadata introduced in Chapter 2.2. The semantics as well as possible values of each of the listed attributes is detailed in Appendix B of this document.

Two attributes of the BDD metamodel are of particular importance. Firstly, the attribute “Distinguishing Characteristics” offers the possibility to include pertinent characteristics with specific values for each integration object being defined. The attribute should not serve as a container for all existing properties (this is accomplished by “Attributes” where users can maintain a list of attributes of the object to be defined with short descriptions, the format and field length of each property) but rather those characteristics that are specific to the integration object being defined. This allows demarcating entities of the BDD more clearly from each other. Secondly, the exact characterization of the dependencies to other integration objects is obtained by the attributes “Synonyms” (i.e. integration objects with similar semantics) and “Relationships”. The latter is used to precisely describe the relationship to associated entities in order to provide structural information. Consequently, these attributes realize the requirements postulated in Chapter 2.3.6 that distinguish the BDD from simple glossaries.

The provision of the BDD infrastructure should begin as early as possible and be completed until the start of the integration object definitions at the latest (see Chapter 3.3.2) so that the BDD content can be administered. It is a nonrecurring business tasks meaning that the implementation has to be realized only once (in the initial setup cycle). An essential point within the implementation is the establishment of different roles so that different user groups using the BDD, such as systems designers, application users, management, data owners or database administrators can have access directly to the entities they require and not to more than necessary.

A second important infrastructural prerequisite, particularly in terms of an easy to operate modification workflow for existing definitions, is the implementation of a change request form that is comfortably accessible (e.g. via the company’s intranet) for all possible metadata users.

<table>
<tr><td colspan="3">Business Task
Define Forms for Change Requests</td></tr>
<tr><td>Output
Change Request Form</td><td>Role(s)
o Chief Data Steward (R)</td><td>Technique(s)
---</td></tr>
<tr><td colspan="3">Description
In order to allow for a flexible adaptation of the defined integration objects a Change Request Form should be designed to enable all possible data users to communicate requests for modification directly to the responsible instances. The request form needs to be easily accessible for all employees at their working place. This allows for the above mentioned integration of checkpoints into IT and data management processes (see Table 3-7).</td></tr>
<tr><td colspan="3">Activities
o Design
o Communicate and Make Request Form Available</td></tr>
</table>

Table 3-10: Description of the business task “Define Forms for Change Requests”

3.4 Documentation Model

A documentation model comprises the totality of the design results elaborated within a method [Gutzwiller 1994, p. 14] as well as the relationships and dependencies, both as regards content and time, between these design results. As most of the results were already discussed during the presentation of the procedure model in the chapter before, the focus in this chapter is to highlight the relationships between them in this section of the working paper. Thereafter, we attached templates for the essential documents (as possible design results) that are helpful tools for the responsible roles to carry out their tasks in a structured way and achieve the intended results and, at the mean time, represent the pivotal results of the different phases.

Table 3-11 summarizes the design results of all business tasks of METIO. These design results do not only represent documents (such as 1.3, 2.1 or 3.1) but can also constitute an implemented tool or a definition. The table shows the relationships between the outcomes signifying which result A is depending and using which result B. The document containing the set of integration objects relevant to be defined for example necessitates a list of identified business and data objects respectively as input for consolidation.

	Result B			
Result A	**1**	**2**	**3**	**4**
1. Identification of Integration Objects				
1.1 Essential BOs (with prioritization)				
1.2 Essential Data Objects (with prioritization)				
1.3 Consolidated set of IOs (with prioritization)	1.1, 1.2			
2. Definition of Integration Objects				
2.1 IO classification and assignment	1.3			
2.2 Definitions for all relevant IOs	1.3	2.1		4.1
3 Integration with Data Management Processes				
3.1 Triggers for metadata refinement		2.1		
4 Provision of Technical Infrastructure				
4.1 BDD Tool infrastructure				
4.2 Change Request Forms				

Table 3-11: Dependencies between the design results of METIO

The document “IO classification and assignment” needs some further specification as it is of particular importance for some of the following business tasks. The document template is illustrated in Figure 3-7. As the graphics shows, this document is used, firstly, to classify an integration object in terms of usage scope as well as security and compliance related characteristics which influences directly the definition process thereafter (section "Classification"). Secondly, it lists all Business and Technical Data Stewards responsible for defining metadata of the integration object (section "Assignment"). This, in turn, has to be considered later on when triggers for metadata refinement are defined that need to be detected by the corresponding roles. In the document the names of the responsible Data Stewards can be registered and their responsibilities can be specified according to the RACI model (see Chapter 3.3).

IO classification and assignment

Name of the Integration Object:			
Classification			
Integration Object used across multiple organizational units?	☐	Number?	☐
Integration Object used in multiple processes?	☐	Number?	☐
Integration Object used by multiple applications?	☐	Number?	☐
Security relevant Integration Object?	☐		
Compliance related Integration Object?	☐		
Assignments			
Business Data Steward		Technical Data Stewards	
	☐		☐
	☐		☐
	☐		☐
	☐		☐
	☐		☐
	☐		☐
	☐		☐
Legend			
R: Responsible A: Accountable C: Consulted I: Informed			

Figure 3-7: Document template for the classification and assignment of integration objects

The document template for defining triggers for metadata refinement (business task 3.1) illustrated in Figure 3-8 contains some examples for possible triggers that might necessitate the adaptation of existing definitions and, consequently, the re-entry in den metadata refinement cycle as indicated in Figure 3-5. The triggers are detailed with potential ways of measuring them and the responsible roles for supervising the measurement. Furthermore, the list should include a threshold that defines a limit value showing when a new iteration of the metadata refinement is required (in order to avoid that every single change request leads to a new iteration). In general, this threshold will vary between companies. However, there will be triggers, such as a change in regulations or the integration of other companies (mergers or acquisitions) that will require an immediate adaptation of existing definitions.

Triggers for metadata refinement

Triggers	Measurement	Threshold	Responsibility
Additional integration object identified	number of change requests concerning data structure of applications	company-specific determination	Technical Data Steward
	number of change requests to integration object owner		Integration Object Owner (Business Data Steward)
Insufficient integration object definition (missing attributes)	number of change requests to integration object owner	company-specific determination	Integration Object Owner (Business Data Steward)
Differing understanding of integration objects	number of requests concerning understanding of definitions	company-specific determination	Integration Object Owner (Business Data Steward)
Merger&Acquisition / Outsourcing	---	immediately	Management
Change in compliance demands	---	immediately	Legal Department, respective Business Units
...			

Figure 3-8: Document template „Triggers for metadata refinement" (with examples)

3.5 Techniques

In this chapter we describe the techniques mentioned in Chapter 3.3 in more detail. They constitute an important component for the responsible roles to achieve the necessary results in each of the defined business tasks. As we mainly refer to well-established techniques (written in italic in Chapter 3.3) that have proven helpful in a number of projects. These techniques will not be described in detail here but, for simplicity reasons, we rather refer to literature where they are described in detail. They are summarized in the following table.

Name of the Technique	Reference	Used in Process Step
Business Process Analysis	[Biazzo 2000], [IMG 1997, pp. TECH 37 ff.]	1.1 Analyze Business Processes
Business Object Analysis	[IMG 1997, pp. TECH 22 ff.]	1.1 Analyze Business Processes
Information/Data Flow Analysis	[DeMarco 1979], [Buckl et al. 2008]	1.2 Analyze Application Architecture
Group/Individual/Expert Interview	[Gläser/Laudel 2006], [Bortz/Döring 2006]	1.1 Analyze Business Processes, 1.2 Analyze Application Architecture
Content Analysis	[Gläser/Laudel 2006], [Myers 2002]	1.1 Analyze Business Processes, 1.2 Analyze Application Architecture
Requirements Analysis	[Hay 2002], [Mylopoulos et al. 1999], [Zmud 1983], [Taggart/Tharp 1977]	4.1 Implement BDD
Software Testing	[Sneed et al. 2006], [Hetzel 1988], [Thaller 2002]	4.1 Implement BDD

Table 3-12: Techniques utilized within METIO

For the conduction of interviews, we developed an interview guideline that is intended to facilitate the interviewee to receive the necessary information for both identification and description of the integration objects. The interview guideline is appended in Appendix D.

4 Application of METIO in a Real-World Setting

For the purpose of establishing a better understanding of the method and of gaining insights for a first evaluation of METIO, we will present adoption of our approach in a real-world scenario, namely at Deutsche Bahn Netz (DB Netz), in the following. The affected area of application was the Infrastructure Data Management (IDM) that is responsible for the long term increase and assurance of the quality of infrastructural data as well as the harmonization of data models and data management processes.

Project Goals The establishment of a Business Data Dictionary at DB Netz was initiated to allow for the integration of Business and IT with regard to a common language. The intended outcomes of the project include:

- A generic metadata model for IDM,
- A "Single Source of Truth" serving as a reference for all employees,
- Defined accountabilities, and
- Analysis and improvement of application interface.

The intention of adopting METIO within DB Netz was to obtain a detailed corporate description and understanding of the company's essential business objects and their attributes. Moreover, both relationships between different entities and dependent attributes should be revealed.

Project Scope Definition The analysis was deliberately constrained to one single process, in our case the reconstruction of a train platform, in order to constrain the initial effort. The rationale for choosing this process was based on the following three criteria:

- The process should span over multiple organizational units (in our case DB Netz, DB Energie, DB Projektbau etc.) and several information systems (as a lack of transparency becomes notably evident beyond functional boundaries);
- The process should possess limited, i.e. still manageable complexity; and
- The process should be well defined.

Thereafter, an extension of the scope of application is intended for the near future.

Business Process Analysis After a department internal inventory of all infrastructure systems and business objects we began a consolidated, cross-department analysis. In several workshop sessions with process owners and line managers from concerned departments (such as DB Station & Service, DB Energie etc.) 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. Due to their relevance for reporting and obtaining federal budgets, we notably focused on entities of the "Infrastrukturkataster", a register in which all infrastructure information is stored. An extract of the resulting information flow chart is depicted in Figure 4-1.

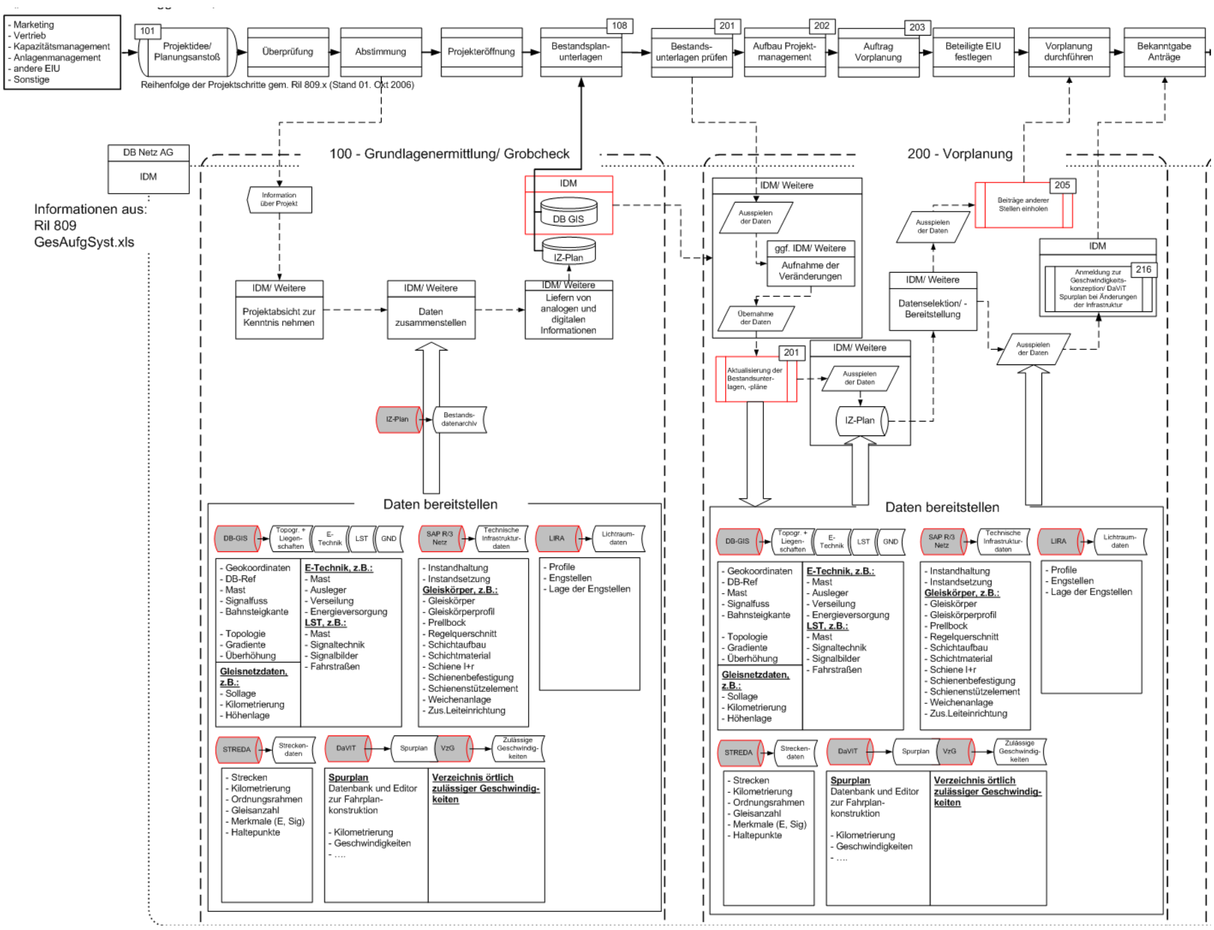

Figure 4-1: Information flow diagram for the business process "Reconstruction of a Train Platform" (extract)

Application Landscape Analysis According to the METIO Procedure Model, 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. This enabled us to get our attention on the discrepancy between the processes and the applications supporting them.

On the system level in a first step the applications that provide data for the different business process segments were revealed. The corresponding application owners then carried out a data flow analysis identifying relevant interfaces and data objects stored within their applications and exchanged between them based on existing system diagrams and data flow maps. Figure 4-2 illustrates the result of the data object identification process with the consolidated list of the most relevant data objects that are used by multiple systems. The list was used as input for the following phase in which the entities were mapped to the identified business objects.

Objekt:	DB-GIS	R/3-K	R/3-Netz	GFD-I	DaViT Spurplan	STREDA S-GIS	AG 850	APS	TPS
Bahnübergang	☑	☑	☑	☑	☑	☑			
Betriebsstellen	☑		☑	☑	☑	☑	☑	☑	☑
Brücken	☑	☑	☑			☑			
Gleis	☑	☑	☑	☑	☑			☑	
Hauptsignal	☑		☑	☑	☑				
KM-Sprünge	☑			☑	☑	☑			
Stellwerke	☑	☑	☑						
Strecken				☑	☑	☑			☑
Tunnel	☑	☑	☑	☑		☑			
Weichen	☑	☑	☑	☑	☑			☑	

Figure 4-2: List of relevant data objects with assignment to corresponding applications

Integration Object Identification Based on the results of the first two tasks, a list of relevant integration objects was compiled in a dedicated consolidation workshop in which business and data objects were mapped to each other. The misfits discovered during this consolidation mapping led to new iterations of either the business process or the application landscape analysis. As a result we were able to obtain a consistent, mutually agreed-upon list of integration objects as input for the definition process. Due to the relatively low number of identified integration objects a prioritization was not considered necessary.

Integration Object Definition With regard to the description of the identified integration 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 integration 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 Business Data Dictionary.

Figure 4-3 shows an extract from the wiki-based BDD at DB Netz showing the initial description structure and definition for a track switch ("Weiche").

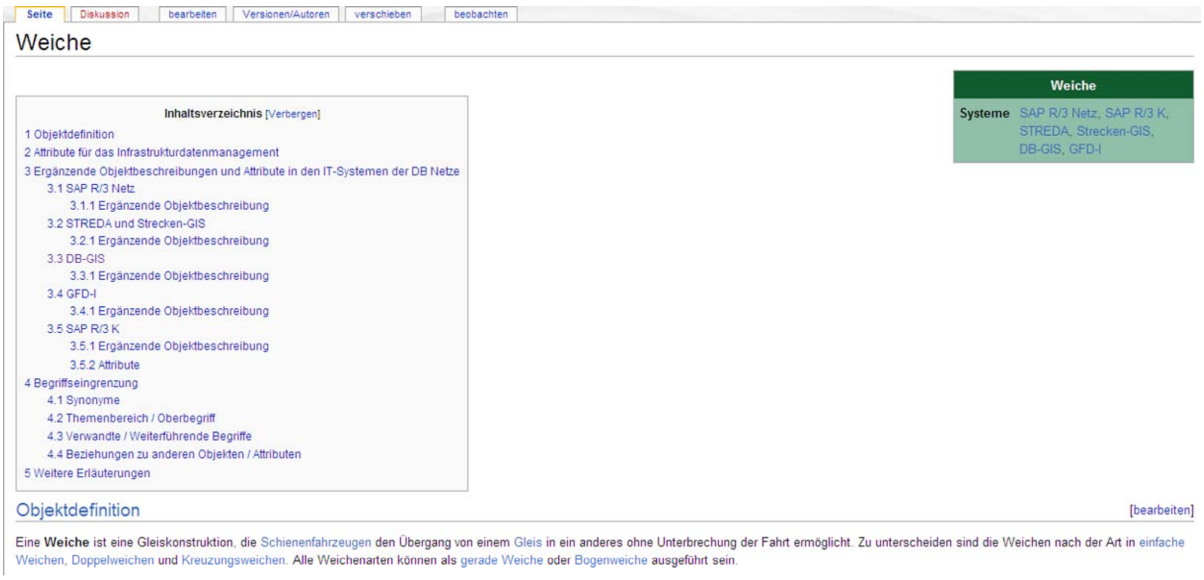

Figure 4-3: Extract from the wiki-based BDD with the entry for a track switch ("Weiche")

5 Summary and Outlook

The working paper at hand introduced a method that is intended to enable companies to increase the transparency and consistency among business and data objects on a conceptual level. Therefore, relevant entities need to be identified and then precisely defined. The (positive) consequences of unambiguously defined integration objects are manifold: Firstly, they ensure a common understanding of important information entities 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 integration 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 [O'Neil 2005a, p. 4]. And lastly, they are a prerequisite for and facilitator of semantic information integration of heterogeneous applications that need to communicate with each other and exchange data [March/Kim 1988, p. 7].

After a short introduction that motivated the topic and outlined its relevance for companies today, we developed the conceptual foundation by first presenting the metamodel of Method Engineering that specifies all components of a method and was, consequently, considered as a framework for METIO. Thereafter, we discussed the significance of metadata as an essential concept for precisely defining integration objects and their attributes. And finally, we introduced and evaluated different concepts for the management of metadata by means of criteria that we derived from our experience within the Competence Center for Corporate Data Quality. As a result of this evaluation we found Business Data Dictionaries as the most appropriate tool to manage data from a business oriented perspective (including the seamless integration of metadata in the employee’s daily work) and for our intended target group – the business user.

Chapter 3 then contains the actual Method for Establishing Transparency on Integration Objects described in detail with all constituents defined by the Method Engineering approach: the metamodel, necessary roles, the procedure model (including the activities necessary to be carried out), the documentation model and the techniques. 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 business and data objects. The following chapter showed application of METIO in a real-world scenario, namely at DB Netz, one of our partner companies. The adoption of the method revealed that already the initial definition process can take a considerable period of time (up to several years) and is not finalized yet. After finalizing the adoption of our method within the organization, 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 or already taking place.

One of the central elements addressed with METIO is the ongoing process that is needed for keeping transparency and consistency high. It is not sufficient to precisely define integration objects one time, but to continuously control and improve metadata once transparency has been established. The Data Governance roles (see Chapter 3.2) and the explicit consideration of possibilities to integrate data and metadata management processes (see Chapter 3.3.3) tried to respond to this requirement and to ensure that the quality of definitions is kept on a high level.

Another aspect that has been touched on in Chapter 3.1 was the utilization of metadata for data quality measurement. We did not broach this topic in detail here as it was not within the scope of the working paper. However, the metamodel incorporates a link to the subject of Data Quality Scorecards that could be extended in one of the following working papers.

As shortly mentioned in the introductory chapter, this working paper with the method it contains is a first important step towards an Information Architecture for Corporate Data Quality. The knowledge about the central business entities of a company, their significance as well as their origins (in processes and applications) represents the prerequisite for analysing the underlying application architecture and the information flows between applications. The Information Architecture, in turn, is the basis for high quality business information and data integration [Schreiber 2003]. The elaboration of concepts for designing and developing an Information Architecture, including information flow and data distribution analysis, will be subject to further research. Moreover, the documentation can serve as helpful input when introducing Data Governance or a company-wide Master Data Management.

In a next step, the elaboration of a metric for evaluating METIO will represent an integral part of our future research. This will allow us to reveal advantages and benefits of the method, but at the same time identify aspects for further refinement and improvement of the proposed approach.

We would like to emphasize the importance of integrating metadata – and consequently the content of the BDD – in other applications that are used by the ordinary employee in order to facilitate his work and improve the quality of his results. MARCO points out the relevance of metadata to be accessible in the easiest way possible for all users in order to reduce their efforts and time to retrieve information, e.g. via a web interface [Marco 2000, pp. 34 f.]. The goal should be to enable business people to capture the largely unstructured data that they uncover every day in order to help them perform their jobs better. So-called hover boxes (i.e. boxes appearing when the user hovers over a certain word) or even select lists are conceivable concepts to integrate metadata in applications and make them easily accessible to users at the moment when they need them. For the purpose of this integration the Competence Center is currently developing a tool that allows seamless desktop integration of the corporate BDD, i.e. access to the corporate BDD directly from the application an employee is working with. This should considerably reduce the search effort for employees in their daily business and, consequently, increase work efficiency.

The Business Data Dictionary (particularly the synonyms and related terms maintained) as tool used for managing and structuring metadata, sets the stage for a "company-wide semantic web". This would allow for intelligent search and real knowledge management in the sense of a corporate, shared knowledge base where different groups within the organization can use existing company knowledge and terminology. An extension to ontology-based metadata and definition management that increases the possibilities of mapping relationships and links between integration objects as well as enable intelligent, semantically enriched search mechanisms might constitute a next step towards an extension of METIO. Concepts such as the Resource Description Framework (RDF) or the Web Ontology Language (OWL) allow related information to be tagged for easy navigation and search. First publications raising the topic of using wiki technology for metadata and knowledge management have been appeared recently [Vrandecic et al. 2005, Krötzsch et al. 2007].

References

[Australian Government Locator Service 2002]
Australian Government Locator Service, AGLS Metadata Element Set – Part1: Reference Description. Version 1.3, National Archives of Australia, Canberra, Australia 2002
[Balzert 2000]
Balzert, H., Lehrbuch der Software-Technik - Software-Entwicklung, Aufl., Spektrum Akademischer Verlag, Heidelberg 2000
[Berson/Dubov 2007]
Berson, A., Dubov, L., Master Data Management and Customer Data Integration for a Global Enterprise, Aufl., McGraw-Hill, 2007
[Biazzo 2000]
Biazzo, S., Approaches to business process analysis: a review, in: Business Process Management, 6, 2000, Nr. 2, S. 99-112
[Bortz/Döring 2006]
Bortz, J., Döring, N., Forschungsmethoden und Evaluation - für Human - und Sozialwissenschaftler, 4., überarbeitete Aufl. Aufl., Springer, Berlin 2006
[Brinkkemper 1996]
Brinkkemper, S., Method Engineering: Engineering of Information Systems Development Methods and Tools, in: Information and Software Technology, 38, 1996, Nr. 4, S. 275-280
[Brunner/Dinter 2003]
Brunner, J., Dinter, B., Vom Data Warehouse zum Business Performance Management - Anforderungen an das Metadatenmanagement, in: von Maur, E., Winter, R. (Hrsg.), Data Warehouse Management, Aufl., Springer, Berlin et al. 2003, S. 291-311
[Buckl et al. 2008]
Buckl, S., Ernst, A. M., Lankes, J., Matthes, F., Enterprise Architecture Management Pattern Catalog, Working Paper, 2008
[Burnett et al. 1999]
Burnett, K., Ng, K. B., Park, S., A Comparison of the Two Traditions of Metadata Development, in: Journal of the American Society for Information Science, 50, 1999, Nr. 13, S. 1209-1217
[Chisholm 2006]
Chisholm, M., Master Data versus Reference Data, in: DMReview.com, 2006, Nr.
[Davenport/Short 1990]
Davenport, T. H., Short, J., The New Industrial Engineering - Information Technology and Business Process Redesign, in: Sloan Management Review, 1990, Nr. Summer, S. 11-27
[DeMarco 1979]
DeMarco, T., Structured Analysis and System Specification, Aufl., Prentice Hall, 1979
[Derungs 1997]
Derungs, M., Workflowsysteme zur Prozessumsetzung, Dissertation, Difo-Druck, Bamberg 1997
[Devlin/Murphy 1988]
Devlin, B. A., Murphy, P. T., An architecture for a business and information system, in: IBM Systems Journal, 27, 1988, Nr. 1, S. 22
[Dublin Core Metadata Initiative 2008]
Dublin Core Metadata Initiative, Dublin Core Metadata Element Set, Version 1.1, http://dublincore.org/documents/dces, 10.03.2008
[Dyché/Levy 2006]
Dyché, J., Levy, E., Customer Data Integration, Aufl., John Wiley & Sons, Hoboken, New Jersey 2006
[Elmasri/Navathe 2000]
Elmasri, R., Navathe, S. B., Fundamentals of Database Systems, 3. Aufl., Addison Wesley, Reading et al. 2000
[English 1999]
English, L. P., Improving Data Warehouse and Business Information Quality, 1. Aufl., John Wiley & Sons, Inc., New York, NY 1999
[Ferstl/Sinz 1995]
Ferstl, O. K., Sinz, E. J., Der Ansatz des Semantsichen Objektmodells (SOM) zur Modellierung von Geschäftsprozessen, in: Wirtschaftsinformatik, 37, 1995, Nr. 3, S. 209-220
[Ferstl/Sinz 2001]
Ferstl, O. K., Sinz, E. J., Grundlagen der Wirtschaftsinformatik, 4. Auflage. Aufl., Oldenbourg, München 2001
[Frank 1994]
Frank, U., Multiperspektivische Unternehmensmodellierung: Theoretischer Hintergrund und Entwurf einer objektorientierten Entwicklungsumgebung, Aufl., Oldenbourg, München 1994
[Friedman 2006]
Friedman, T., Gartner Study on Data Quality Shows That IT Still Bears the Burden, Working Paper, 2006
[Gläser/Laudel 2006]
Gläser, J., Laudel, G., Experteninterviews und qualitative Inhaltsanalyse, 2. Aufl., VS Verlag fÃ¼r Sozialwissenschaften, Wiesbaden 2006

[Goldkuhl et al. 1997]
Goldkuhl, G., Lind, M., Seigerroth, U., Method Integration as a Learning Process, Proceedings of the Fifth International Conference of the British Computer Society Information Systems Methodologies Specialist Group, Springer, 1997, S. 15-26
[Greiffenberg 2003]
Greiffenberg, S., Methoden als Theorien der Wirtschaftsinformatik, Wirtschaftsinformatik 2003 Band II - Medien, Märkte, Mobilität, Heidelberg, Physica-Verlag, II, 2003, S. 947-968
[Gruber 1993]
Gruber, T. R., A Translation Approach to Portable Ontology Specifications, in: Knowledge Acquisition, 5, 1993, Nr. 2
[Gutzwiller 1994]
Gutzwiller, T., Das CC RIM-Referenzmodell für den Entwurf von betrieblichen, transaktionsorientierten Informationssystemen, Aufl., Physica, Heidelberg 1994
[Hay 2002]
Hay, D. C., Requirements Analysis: From Business Views to Architecture, Aufl., Prentice Hall, Upper Saddle River 2002
[Hetzel 1988]
Hetzel, B., The complete guide to software testing, 2. Auflage. Aufl., QED, Wellesley 1988
[Heym 1993]
Heym, M., Methoden-Engineering - Spezifikation und Integration von Entwicklungsmethoden für Informationssysteme, Dissertation, Hochschule St. Gallen, Hallstadt 1993
[IMG 1997]
IMG, PROMET BPR, Methodenhandbuch für den Entwurf von Geschäftsprozessen, Version 2.0, Aufl., Information Management Group/Institut für Wirtschaftsinformatik Universität St. Gallen, St. Gallen 1997
[ISO 1995]
ISO, Information Technology – Metadata Registries (MDR). Part 4: Formulation of Data Definitions, International Organization for Standardization (ISO), Geneva, Switzerland 1995
[ISO 2003]
ISO, Information Technology – Metadata Registries (MDR). Part 3: Registry Metamodel and Basic Attributes, International Organization for Standardization (ISO), 2003
[IT Governance Institute 2005]
IT Governance Institute, CobiT 4.0: Control Objectives, Management Guidelines, Maturity Models, IT Governance Institute, Rolling Meadows/IL 2005
[Krallmann 1996]
Krallmann, H., Systemanalyse im Unternehmen: Geschäftsprozessoptimierung, partizipative Vorgehensmodelle, objektorientierte Analyse, Zweite Auflage. Aufl., Oldenbourg, München 1996
[Krötzsch et al. 2007]
Krötzsch, M., Vrandecic, D., Völkel, M., Haller, H., Studer, R., Semantic Wikipedia, in: Journal of Web Semantics, 5, 2007, Nr. 4, S. 251--261
[Lankhorst 2005]
Lankhorst, M., Enterprise Architecture at Work: Modelling, Communication and Analysis, Aufl., Springer, Berlin et al. 2005
[Leist/Zellner 2006]
Leist, S., Zellner, G., Evaluation of Current Architecture Frameworks, Proceedings of the 21st Annual ACM Symposium on Applied Computing (SAC2006), April 23-27, 2006, New York, 03.04.2006, ACM Press, 2006, S. 1546-1553
[Leser/Naumann 2007]
Leser, U., Naumann, F., Informationsintegration - Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen, Aufl., dpunkt.verlag, Heidelberg 2007
[March/Kim 1988]
March, S. T., Kim, Y.-G., Information Resource Management: A Metadata Perspective, in: Journal of Management Information Systems, 5, 1988, Nr. 3, S. 5-18
[Marco 2000]
Marco, D., Building and Managing the Meta Data Repository. A Full Lifecycle Guide, Aufl., John Wiley & Sons, New Jersey 2000
[Marco/Smith 2006]
Marco, D., Smith, A. M., Metadata Management & Enterprise Architecture: Understanding Data Governance and Stewardship, in: DM Review, 2006, Nr. Sep/Okt/Nov
[Morschheuser et al. 1996]
Morschheuser, S., Raufer, H., Wargitsch, C., Challenges and Solutions of Document and Workflow Management in a Manufacturing Enterprise: A Case Study, Proceedings of the Hawaii Conference on System Sciences (HICSS '96), Los Alamitos, IEEE Computer Society Press, 1996
[Myers 2002]
Myers, M. D., Qualitative Research in Information Systems, http://www.qual.auckland.ac.nz/index.htm, 16.09.2002

[Mylopoulos et al. 1999]
Mylopoulos, J., Chung, L., Yu, E., From Object-Oriented to Goal Oriented Requirements Analysis, in: Communications of the ACM, 42, 1999, Nr. 1, S. 31-37

[Newman/Logan 2006]
Newman, D., Logan, D., Achieving Agility: How Enterprise Information Management Overcomes Information Silos, Gartner Research, Stamford, CT 2006

[NISO 2004]
NISO, Understanding Metadata, NISO (National Information Standards Organization) Press, Bethesda, MD 2004

[Nuseibeh et al. 1996]
Nuseibeh, B. A., Finkelstein, A., Kramer, J., Method Engineering for Multi-Perspective Software Development, in: Information and Software Technology, 38, 1996, Nr. 4, S. 267-274

[O'Neil 2005a]
O'Neil, B. K., Business Metadata: How To Write Definitions, http://www.tdan.com/i032fe01.htm, 27.08.2007

[O'Neil 2005b]
O'Neil, B. K., Launching a Corporate Glossary, http://www.b-eye-network.com/view/1014, 28.08.2007

[Obrst 2003]
Obrst, L., Ontologies for Semantically Interoperable Systems, Proceedings of the 2006 international workshop on Global integrated model management, New York, ACM Press, 2003, S. 366-369

[Olson 2003]
Olson, J., Data Quality - The Accuracy Dimension, Aufl., Morgan Kaufmann, San Francisco 2003

[OMG 1996]
OMG, Common Facilities RFP-4: Common Business Objects and Business Object Facility, Object Management Group (OMG), Framingham, MA 1996

[OMG 2003]
OMG, MDA Guide - Version 1.0.1, 2003

[Österle 1995]
Österle, H., Business Engineering: Prozess- und Systementwicklung, Band 1: Entwurfstechniken, 2. Aufl., Springer, Berlin 1995

[Österle 2004]
Österle, H., Übergang zur Informationsgesellschaft (New Economy), in: Dubs, R., Euler, D., Rüegg-Stürm, J., Wyss, C. (Hrsg.), Einführung in die Managementlehre, Band 1, Aufl., Haupt, Bern 2004, S. 507-525

[Päivärinta et al. 2002]
Päivärinta, T., Tyrväinen, P., Ylimäki, T., Defining Organizational Document Metadata: A Case Beyond Standards, Proceedings of 10th European Conference on Information Systems (ECIS), Gdansk (Poland), 2002

[Pierce 2005]
Pierce, E. M., Introduction, in: Wang, R.Y., Pierce, E.M., Madnick, S.E. (Hrsg.), Information Quality, Aufl., M.E. Sharpe, 2005, S. 3-17

[Puschmann 2004]
Puschmann, T., Prozessportale - Architektur zur Vernetzung mit Kunden und Lieferanten, Aufl., Springer, Berlin 2004

[Quantz/Wichmann 2003]
Quantz, J., Wichmann, T., E-Business-Standards in Deutschland, Berlecon Research, Berlin 2003

[Rosemann/zur Mühlen 1997]
Rosemann, M., zur Mühlen, M., Modellierung der Aufbauorganisation in Workflow-Management-Systemen: Kritische Bestandsaufnahme und Gestaltungsvorschläge, Proceedings of the EMISA-Fachgruppentreffen, 1997, S. 78-84

[Rosenthal et al. 2004]
Rosenthal, A., Seligman, L., Renner, S., From Semantic Integration to Semantics Management: Case Studies and a Way Forward, in: ACM SIGMOD Record, 33, 2004, Nr. 4, S. 44-50

[Rüegg-Stürm 2002]
Rüegg-Stürm, J., Das neue St. Galler Management-Modell: Grundkategorien einer integrierten Managementlehre - Der HSG-Ansatz, Aufl., Haupt, Bern, Stuttgart, Wien 2002

[Rumbaugh 1995]
Rumbaugh, J., What is a method?, in: Journal of Object Oriented Programming, 8, 1995, Nr. 6, S. 10-26

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

[Schreiber 2003]
Schreiber, Z., Semantic Information Architecture: Creating Value by Understanding Data, http://www.dmreview.com/article, 10.06.2005

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

[Shankaranarayanan/Even 2004]
Shankaranarayanan, G., Even, A., Managing Metadata in Data Warehouses: Pitfalls and Possibilities, in: Communications of AIS, 14, 2004, Nr., S. 247-274

[Shankaranarayanan/Even 2006]
Shankaranarayanan, G., Even, A., The Metadata Enigma, in: Communications of the ACM, 49, 2006, Nr. 2, S. 88-94

[Shankaranarayanan et al. 2003]
Shankaranarayanan, G., Ziad, M., Wang, R. Y., Managing Data Quality in Dynamic Decision Environments: An Information Product Approach, in: Journal of Database Management, 14, 2003, Nr. 4, S. 14-32

[Sneed et al. 2006]
Sneed, H., Baumgartner, M., Seidl, R., Der Systemtest - Anforderungsbasiertes Testen von Software-Systemen, Aufl., Carl Hanser, Munich/Vienna 2006

[Sommerville 2007]
Sommerville, I., Software Engineering, 8. Aufl., Pearson Studium, 2007

[Strahringer 1996]
Strahringer, S., Metamodellierung als Instrument des Methodenvergleichs, Aufl., Shaker, Aachen 1996

[Taggart/Tharp 1977]
Taggart, W. M., Tharp, M. O., A Survey of Information Requirements Analysis Techniques, in: Computing Surveys, 9, 1977, Nr. 4

[Thaller 2002]
Thaller, G. E., Software-Test. Verifikation und Validation, Aufl., Heise, Hannover 2002

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

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

[Uschold/Gruninger 2004]
Uschold, M., Gruninger, M., Ontologies and Semantics for Seamless Connectivity, in: SIGMOD Record, 33, 2004, Nr. 4, S. 58-64

[von Halle et al. 2006]
von Halle, B., Goldberg, L., Zachman, J., Business Rule Revolution: Running Business the Right Way, Aufl., Happy About, Silicon Valley 2006

[Vrandecic et al. 2005]
Vrandecic, D., Pinto, S., Tempich, C., Sure, Y., The DILIGENT Knowledge Processes, in: Journal of Knowledge Management, 9, 2005, Nr. 5, S. 85--96

[Wang/Strong 1996]
Wang, R. Y., Strong, D. M., Beyond Accuracy: What Data Quality Means to Data Consumers, in: Journal of Management Information Systems, 12, 1996, Nr. 4, S. 5-34

[Zmud 1983]
Zmud, R. W., Information Systems in Organizations, Aufl., Scott, Foresman and Company, Glenview, IL 1983

Appendix A: Glossary

This glossary provides an overview over the meta entity types of the METIO metamodel (left column of Table 0-1) with their concrete, semantically precise definitions and their respective relationships to other meta entity types (right column).

Metadata Definition	
Business Data Dictionary (BDD)	A BDD stores a set of metadata that contains semantically precise definitions (terminological metadata) for each integration object, its structure and relationship to other entities (structural metadata), the accountabilities for the integration object (governance metadata) as well as compliance related information (compliance metadata) and where it is used (usage metadata) with the goal of increasing enterprise-wide transparency. Relationships: A BDD stores a set of metadata. It is the essential output of the metadata definition and maintenance process and its context is accessed by data usage processes.
Metadata	Metadata represents structured data that describes other data (its meaning and characteristics) [Burnett et al. 1999, p. 1212]. Metadata constitutes a “means by which the *structure* and *behaviour* of data is recorded, controlled, and published across an organization” [Tozer 1999, p. xix]. Relationships: Metadata is stored in at least one BDD. Metadata can describe characteristics of one to many integration objects as well business objects. Metadata can be subdivided into administrative metadata, terminological metadata (such as definitions), structural metadata, governance metadata, compliance metadata and usage metadata.
Administrative Metadata	Administrative metadata provides general information about an entity, such as its priority, its current status, its definition process, general comments, its maintenance and instancing process etc. [NISO 2004, p. 1]. Relationships: Administrative metadata is a specification of metadata in general.
Structural Metadata	Structural metadata of an integration object is metadata describing the structure of an entity (such as necessary attributes) as well as relationships to other integration objects [NISO 2004, p. 1].

	Relationships: Structural metadata is a specification of metadata in general.
Terminological Metadata	Terminological metadata of an integration object is metadata describing the meaning (semantics) of an entity through a unique name, its synonyms and an unambiguous definition. Relationships: Terminological metadata is a specification of metadata in general.
Usage Metadata	Usage metadata of an integration object is metadata revealing how the entity is embedded, i.e. in which applications, processes and organizational units the integration object is used. Consequently, attributes comprise amongst others “Usage in Processes” and “Usage in Applications”. These attributes are particularly important for information architecture analysis and design (e.g. information flows). Relationships: Usage metadata is a specification of metadata in general.
Governance Metadata	Governance metadata contain all metadata related to responsibilities/accountabilities for integration objects. Relationships: Governance metadata is a specification of metadata in general.
Compliance Metadata	Compliance metadata consists of metadata resulting from regulations influencing or constraining the behavior and use of the integration object, but also comprises business rules, i.e. general rules or constraints that need to be considered for the sake of consistency when using the integration object [Tozer 1999, p. 10, von Halle et al. 2006, p. 6]. Moreover, we integrate security related metadata due to their close relation to compliance issues. Relationships: Compliance metadata is a specification of metadata in general.
Integration Object	Integration objects are abstract entities that are used only for consolidation purposes, namely to enable a consistent mapping of process-specific business objects and application-specific data objects on a single unambiguous reference entity. These reference objects need to be semantically described with their entire set of attributes according to the BDD data model developed in this working paper. Hence, integration objects constitute a collection of persistently saved information which is accessed repeatedly. In contrast to

	a business object that is the equivalent on the process level and a data object (or element) as a correspondent on the system level, integration objects represent constructs on a logical level across applications integrating characteristics from both business and data objects. Relationships: An integration object is described by metadata and in general represented by several (but at least one) data objects on the system level. An integration object is always related to at least one (generally to several) other business object(s) on a process level.
Relationship	A relationship is the directed connection between two integration objects. Relationships serve to clarify how an integration object is related to another entity.
Data Quality	
Data Quality Value	A data quality value is the measured value for a data quality metrics. By comparing it with a corresponding target value (for the same metrics) it can serve for target-performance comparison in order to reveal gaps (data quality variance) and indicate the currently achieved quality level. Within the METIO method the actual value is measured to determine the degree of fulfillment of the data and business object instances with the integration objects defined in the BDD in order to evaluate the quality of the metadata definition and maintenance process. Relationships: One up to many data quality values relate to one data quality metric and one corresponding data quality target value as they can be measured more than once.
Data Quality Target Value	A data quality target value is the predefined level that a company aims to achieve with regard to a certain data quality metric. In conjunction with the measured actual values it can serve for target-performance comparison in order to reveal gaps (data quality variance) and indicate the currently achieved quality level. Relationships: At least one up to many data quality target values can be defined for one data quality metric. A target value is contrasted to one up to many data quality values.
Data Quality Variance	A data quality variance is the gap between a predefined data quality target value and one or more measured data quality values/results serving for the analysis of the

	degree of target achievement. It is a control activity which is part of the management level and which is essential for continuous analysis and improvement process. Relationships: Multiple data quality variances (with respect to several metrics) can be conducted for (and, in case of a gap between target and actual value, trigger) a metadata definition and maintenance process.
Data Quality Metric	A data quality metric is a mean to quantify and measure the quality of data within a company. A metrics can be either an absolute measure or a ratio. Relationships: At least one up to many data quality metrics operationalize a data quality dimension. For a data quality metric at least one target value is defined and one up to many actual values are measured.
Data Quality Dimension	A data quality dimension represents a set of data quality attributes relevant for the evaluation and measurement of data quality. They represent a single aspect or construct of data quality [Wang/Strong 1996, p. 6]. Relationships: One up to many data quality dimensions are used within a data quality definition. The dimensions are operationalized by at least one up to many data quality metrics.
Data Quality Definition	A data quality definition contains a number of data quality dimensions through which data quality is conceptualized. It describes what data quality is and which constituent parts it consists of. Relationships: A data quality definition contains at least one up to many data quality dimensions.
Operational and Organizational Structure	
Activity	An activity is a self-contained logical entity in a work flow that – in contrast to a business task – does not generate an economically meaningful output. It subsumes elementary working steps functionally and chronologically and is executed as a whole [cf. Morschheuser et al. 1996]. Relationships: At least one up to many activities constitute the integral component(s) of a business task.
Business Task	A business task is the smallest logical entity within a business process that still generates an economically meaningful output. Business tasks can be performed by

	machines and/or human agents [Österle 1995, p. 50, Rüegg-Stürm 2002, p. 67]. Relationships: A business task consists of at least one up to many activities and is itself integral part of a business process. A business task is executed by exactly one role and generates at least one up to many outputs. At the same time, each business task can use several outputs from other business tasks.
Output	An output is the result of a business process serving, in turn, as input for other business processes. Output can be either material (in case of a product) or immaterial (in case of a service) [Österle 1995, p. 52]. Relationships: An output is generated or used by at least one up to many business processes. The output generated by METIO is a BDD with the metadata for the essential integration objects.
Business Object	A business object is a set of entities with common characteristics and common behaviour representing well-defined business semantics on a process level. This set is generally accepted in the business world (from an international standard or industry best practices). It is a persistent abstract or real-world object that should be able to exist on its own and that is essential for a company's business operations (such as a purchase order or a customer). The business object is an entity description from a business perspective neglecting any implementation specific considerations. As a representation of a thing (such as a person, a business process or concept) active in the business domain, it should at least include its business name and definition, attributes, relationships and rules [OMG 1996, p. 19]. Relationships: One to many business objects (on the process level) are represented by exactly one integration object (on the conceptual level). The business object is a constituent part of one up to many business tasks.
Business Process	A business process subsumes a set of logically related tasks that need to be completed in a certain predetermined order and are supported by IT [Derungs 1997, p. 112]. The business process aims at providing an output relevant for other processes either inside or outside a company by transforming a given input using resources and considering given targets [Davenport/Short 1990, p. 12]. The effectiveness of a process is determined by the specification of the output,

	while its efficiency is constituted by the set targets. Relationships: A business process consists of at least one up to many business tasks and has a predecessor and a successor business process. Moreover, business processes can be nested, i.e. they can contain other business processes. A business process within METIO can be either a Metadata Usage Process, a Metadata Maintenance Process, a Data Usage Process or a Data Maintenance Process.
Application	An application subsumes IT functionality that is closely coupled to each other on a logical level. Relationships: An application uses one to many data objects and supports one to many activities in a business process
Data Usage	An application can use data objects in different ways: create them, read them, update them or delete them. Hence, the access of an application on a data object is put into effect via the data usage entity.
Data Object	A data object – also referred to as a data element – is a coherent, self-contained piece of information suited for automated processing [Lankhorst 2005, p. 99], i.e. the representation of an integration object instance on a system level. It constitutes a category of permanently stored data used by software components [Schwinn 2005, p. 23]. In contrast to an integration object, a data object is always restricted to a single application. Relationships: One or many data objects are used to represent the instance of an integration object.
Metadata Usage Process	A metadata usage process describes the operation of accessing and using metadata that is stored in a metadata repository (like the BDD). It enables users to obtain information about entities they need for their work. Relationships: The metadata usage process represents a specific business process. The process can be accessed by one up to many data usage and maintenance processes.
Metadata Definition and Maintenance Process	The metadata definition and maintenance process provides the basis for establishing transparency on business and data objects in a company. The process is initiated periodically through triggers (such as an unsatisfactory data quality variance) in order to keep the transparency and consistency on a high level.

	Relationships: The metadata definition and maintenance process represents a specific business process. It can be triggered by a data quality variance.
Data Usage Process	A data usage process represents the operation of accessing, adding, deleting, using, changing/updating and saving data objects during the operational work. Relationships: The data usage process represents a specific business process. During the process it can interact with multiple metadata usage processes by accessing metadata information (and one up to many data usage processes can access a metadata usage process).
Data Maintenance Process	A data maintenance process comprises all activities carried out in order to administer and clean the existing data stock. Data maintenance is an ongoing process that can be done either manually or with the help of automated programs. Relationships: The data maintenance process represents a specific business process. The data maintenance process can interact with multiple metadata usage processes by accessing metadata information (and one up to many data maintenance processes can access a metadata usage process).
Role	A role bundles different business tasks that are carried out by a single person (employee) or an organizational unit as well as their area of responsibility and competencies [Österle 2004, p. 515, Puschmann 2004, p. 90]. Relationships: Roles are fulfilled by at least one up to many employees and/or organizational units. Roles can also contain other roles and positions and execute at least one up to many business tasks.
Position	A position is the smallest structural unit within an organization to which certain business tasks and resources are assigned [Österle 1995, p. 98]. Relationships: A position can be fulfilled by multiple (but at least one) employees and can contain several roles. Each position is assigned to exactly one organizational unit.
Employee	An employee is a human agent fulfilling at least one position in a company [Rosemann/zur Mühlen 1997, p. 102] that is employed by an employer and remunerated for his work.

	Relationships: An employee is assigned to a position. Each employee is part of exactly one organizational unit and can fulfill different (one to many) roles.
Organizational Unit	An organizational unit is the aggregation of one or more positions to a self-contained, independent part of the organizational structure [Österle 1995, p. 51]. The relationships between organizational units compose the organizational structure of a company [Rosemann/zur Mühlen 1997, p. 103]. Relationships: An organizational unit consists of at least one up to many positions and employees.

Table 0-1: Definition of the METIO meta entity types

Appendix B: Description of Attributes and Relationships of the BDD

Each of the attributes of the BDD to be implemented is described in the following table with its name, semantics (definition) and possible values.

Terminological Metadata		
Name	name of the entity being defined	
Synonyms	other names used to describe the concept being defined	Multiple items divided by semicolon
Keywords	words or phrases important for searching, accessing, or retrieving information from this concept or data definition	Multiple items divided by semicolon
Definition	provides a verbal definition of the concept	
Potential Values	provides list of the potential values of the concept / attributes	
Coding and Descriptive Convention	Certain concepts may have an identifier or code used to reference each instance of the concept and/or one or more descriptions. This section is to itemise such descriptions together with any format and content conventions that the business has agreed on. The advantage of adding this here is that it enables the reader to have a better view of what the concept will actually be identified by and it avoids having inhomogeneous definitions for each description.	

Structural Metadata		
Relationship (with Type and Direction)	list of terms that have relevance to the concept being defined but are not synonymous	Type of relationship is specified from predefined list of bidirectional relations (see Table 0-3). Terms defined as related terms are automatically updated by a vice verse relationship.
Attributes	List of attributes of the object with a short description for each attribute as well as a specification of its cardinality, format and field length in applications.	Attribute Name │ Short description │ Cardinality │ Attribute Type/Format │ Field length Ex.: Name │ Name of the product │ 1 │ Text/String │ 20
Distinguishing Characteristics	pertinent attributes with specific values for the concept being defined; particularly in order to distinguish from → Related Terms	
Administrative Metadata		
Definition Progress	attribute to track the progress of the definition process for an integration object. depending on the kind of definition process the integration object runs through, the specifications can differ	→ not yet implemented in the BDD Tool
Rationale for	documents the basis or underlying	

Definition	reason behind the definition; avoids revisiting the same arguments about the definition again and again	
Internal Definition Process	defines the kind of definition process (see Figure 3-4) the integration object needs to run through for a comprehensive specification Link to → Workflow Information	
Status	current status of the definition If *Status* = *Replaced*, a → Related Term with relationship *replaces/replaced by* must be defined	Draft ¦ Draft in Project Review ¦ Draft in Final Review ¦ Final ¦ Replaced ¦ Retired
Business Rules	Any general rules or principles that constrain the application of the integration object and therefore need to be considered	
Priority	Defines the priority for agreeing upon the concept definition	High ¦ middle ¦ low or based on the SNIA classification (see Chapter 3.3.1)
Comment	any other general comments that do not fit anywhere else	
Usage Metadata		
Subject / Process Area	High level grouping of concepts according to their business domain and/or the business processes for which the integration object is relevant	
Validity Scope	defines the scope of application within	Group ¦ Company ¦ Project

	the company / group; therefore defines, whether the definition is to be agreed group-wide, company-wide or just project-wide → Where Used	
Where Used	list of all projects (*if scope* = *project*) or companies (*if scope* = *company*) that use the concept defined Based on → Scope	
Usage in Processes	Contains a list of processes and/or sub-processes in which the entity is used with the corresponding activity (i.e. created, read, updated or deleted)	Process Name ┃ [Create ¦ Read ¦ Update ¦ Delete]
Usage in Applications	Contains a list of applications in which the entity is used with the corresponding activity (i.e. created, read, updated or deleted)	Application Name ┃ [Create ¦ Read ¦ Update ¦ Delete]
Compliance Metadata		
Security Classification for Definitions	defines the security level for the concept defined, more precisely for the actual definition	Public ¦ Internal Use Only ¦ Confidential ¦ Secret ¦ Hazardous ¦ Unclassified
Security Classification for Instances	defines the security level for concrete the instances of the integration objects. This attribute has to be distinguished from the one before as different security classifications for the definitions and the instances are possible	Unclassified ¦ Internal Use Only ¦ Restricted ¦ Confidential
Relevant Regulations	contains a list of legal terms and directives that are relevant for the	

	integration object	
Maintenance and Instancing		
Maintenance Procedure	defines whether the definition is maintainable or not; in case it is, a → Maintenance Process Documentation should be added	Maintainable ¦ Configured ¦ Derived ¦ Not Maintainable
Maintenance Process Documentation	describes the process with corresponding tasks and roles for maintaining integration objects	
Instancing Process Documentation	describes the process with corresponding tasks and roles for creating integration object instances	
Configuration Requirements		
Strategic Source	system name which is the strategic source of the data	
Standard Name	Standard name for system implementation (e.g. field name)	
Data Type and Field Length	Specifies the data type and field length for system implementation	Plausibility check with → Coding and Descriptive Convention desirable
Required for Setup	specifies whether the concept is relevant for system rollout	
Governance Metadata (Ownership)		
Responsible Business Data Steward	employee responsible for the definition of the integration object and its maintenance (corresponds to the integration object owner). The Business Data Steward is the contact	

	person in case of change requests.	
[Can be extended based on specific needs]		

Table 0-2: Possible Attributes for a BDD implementation

The semantics of possible relationships are listed in Table 0-3.

Direction 1	**Direction 2**
is generalization of (broader term)	is specialization of (narrower term)
replaces	replaced by
consists of	is part of
has	belongs to
miscellaneous	miscellaneous
[can be extended based on specific needs]	

Table 0-3: Semantics of possible Relationships between BDD Entities

Appendix C: Participants of the CC CDQ Workshops

As stated in the text, we used multiple workshops with our competence center project partner (subject matter experts from different companies), to identify and discuss requirements on a metadata repository based on their specific company needs. Both multilateral CC CDQ workshops (where all experts participated and topics were dealt in group works and plenary discussion) and bilateral workshops with each of the project partner were held. This chapter includes a list of all participants with their affiliation to the diverse companies.

Ulrich Kalex	Alfabet	Germany
Klaus Straub	Audi	Germany
Berthold Brauer	Bayer CropScience	Germany
Albert Hatz	Bosch	Germany
Tilo Nirk	Daimler	Germany
Julia Klopfer	Daimler	Germany
Hans Jacoby	DB Netz	Germany
Detlev Glowinski	DB Netz	Germany
Axel Grewe	Deutsche Telekom	Germany
Marco Glass	Deutsche Telekom	Germany
Martin Dreher	E.ON	Germany
Fritz Bütikofer	ETA	Switzerland
Michael Karper	ETA	Switzerland
Oliver Amsler	ETA	Switzerland
Mike Silberman	IBM	Germany
Luis Pereira	ZF	Germany
Heiner Lehnen	ZF	Germany

We would like to thank everybody for their contribution during the workshops.

Appendix D: Interview Guideline with Description Form

1. What department can the integration object be allocated to? (Multiple answers possible)

   i) Acquisition / Purchase
   ii) Production
   iii) Sales
   iv) Employee
   v) Finance
   vi) Others

2. What is the organizational scope of the integration object?

   i) Department
   ii) Business Unit / Business Area
   iii) Company
   iv) Group

3. What domain(s) respectively sub-domain(s) can the integration object be allocated to?

4. Does an cross-organizational (e.g. industry-wide) recognized definition for this integration object exist, e.g. announced by a standardization organization?

   If yes,
   i) What is the definition?
   ii) Is this definition provided by an internal or external source?

   If no,
   How would you define or describe the integration object?

5. Which attributes describe the integration object? (e.g.: entry masks when creating new objects)

   Which of these attributes or characteristics are unique features which therefore allow the classification of the integration object?

   E.g.: The integration object „customer order“ can have attributes like „customer order number“, „model“, „delivery date“, „production number“, „billing date“ etc., whereas the attribute „customer order number“ is unique for the integration object (is not covered by other integration objects).

6. What are possible values of the integration object respectively for its attributes?

7. Does the integration object consist of other objects (sub-entities) (relationship of aggregation or composition)?

   If yes,
   i) What are these sub-entities?

ii) Is this sub-entity “viable” alone?

8. Is the responsibility for this integration object defined?

   If yes,
   i) Who is the integration object owner?
   ii) Who is responsible for the integration object maintenance?

9. Do you know about other views on this definition within the department, the company or group or is the definition used differently from your understanding?

   If yes,
   i) What are they?
   ii) In which departments do they exist?

10. Do other definitions exist, which describe the same concept (synonyms)?

11. Do any established and standardized abbreviations for the information concept exist in the company (being standardized company- or industry-wide)?

12. Do any legal terms and directives exist that are relevant for the integration object?

13. Are any security-relevant limitations (for access) concerning the integration object or its definition necessary?

14. Is the integration object produced by an application in your area of responsibility?
    If not, where do your applications get the object from?

15. In which applications is this integration object in your area of responsibility administrated and how do you access it?

   i) Create
   ii) Read
   iii) Update
   iv) Delete

   Is there a consistent, cross-application unique identifier on a system level?

16. Is there a responsibility for the integration object defined on a system level (e.g. data responsibility)

   If yes,
   i) Who is the owner of the integration object?
   ii) Which tasks arise from this responsibility?

17. Any other relevant information?

Name:

Subject / Process Area:
(for which integration object is relevant)

Usage Scope:
(projects, departments,...)

Related Business Domain(s) or Sub-Domain(s):

Usage in Process:

Definition:

Integration object Attributes:
(Identification of unique features)

Potential Values:

Responsible Role / Owner:
(if existing)

Synonyms:

Related Terms:

Keywords:

Coding and Descriptive Conventions:

Relevant Regulations:

Security Classification:
(for definitions and object entities)

Internal Definition Process:

Status:

Priority:

Usage in Application:

Primary Source:

System Name:

Data Type:

Field Length:

Other Comments: