ORIGINAL ARTICLE

Toward a functional reference model for master data quality management

Boris Otto · Kai M. Hüner · Hubert Österle

Abstract The quality of master data has become an issue of increasing prominence in companies. One reason for that is the growing number of regulatory and legal provisions companies need to comply with. Another reason is the growing importance of information systems supporting decision-making, requiring master data that is up-to-date, accurate and complete. While improving and maintaining master data quality is an organizational task that cannot be encountered by simply implementing a suitable software system, system support is mandatory in order to be able to meet challenges efficiently and make for good results. This paper describes the design process toward a functional reference model for master data quality management (MDQM). The model design process spanned several iterations comprising multiple design and evaluation cycles, including the model's application in a participative case study at consumer goods manufacturer Beiersdorf. Practitioners may use the reference model as an instrument for the analysis, design and implementation of a company's MDQM system landscape. Moreover, the reference model facilitates evaluation of software systems and supports company-internal and external communication. From a scientific perspective, the reference model is a design artifact; hence it represents a theory for designing information systems in the area of MDQM.

Keywords Data quality management · Master data management · Design science research · Reference model

B. Otto (✉) · K. M. Hüner · H. Österle
Institute of Information Management, University of St. Gallen,
Müller-Friedberg-Str. 8, 9000 St. Gallen, Switzerland
e-mail: boris.otto@unisg.ch

🖄 Springer

1 Introduction

1.1 Motivation and problem statement

Master data describes the essential business entities of a company, such as suppliers, customers, products, employees, and assets (Smith and McKeen 2008; Dreibelbis et al. 2008; Loshin 2008). The quality of master data has become an issue of increased prominence in companies. Reasons for that are an increasing number of regulatory and legal provisions companies need to comply with, the need to have a *single source of truth* for company-wide reporting (McCann 2010; Yen 2004), and the demand for having a *360°-view on the customer* (Leser and Naumann 2007; Pula et al. 2003).

An example for the latter is the shift from long-term fixed-line contracts toward a consumer-centric, information services providing business model in the telecommunications industry. Telecommunications providers are forced to manage large amounts of ordering, complaints, contract, service quality, and billing data around individual customer information, which requires significant changes to the way lifecycles and architectures of customer master data are designed today. Consulting company Deloitte summarizes the importance of master data: "Data ascends from the basement to the board room" (Deloitte 2009).

To meet such requirements, master data of good quality (i.e. master data that is up-to-date, consistent and complete) is needed. Companies respond to these requirements by establishing data governance (Weber et al. 2009; Khatri and Brown 2010) or master data management (MDM) (Smith and McKeen 2008; Loshin 2008). Of course, improving and maintaining master data quality is an organizational task that cannot be encountered by simply implementing a software system (Khatri and Brown 2010; English 1999). However, appropriate system support is mandatory in order to be able to meet challenges regarding master data quality efficiently and make for good results.

The consolidation of the market for data quality management (DQM) and MDM systems that is currently under way has been aggravating decision-making regarding the choice of the right product. IBM acquiring Ascential (IBM 2005) and SAP acquiring Business Objects (SAP 2007b) are two examples of this development. Integration of the products acquired with the existing product portfolio in order to be able to offer an integrated solution for quality oriented MDM (in the following *master data quality management*, MDQM) poses new challenges not just for vendors of MDQM systems but also for user companies:

- *System evaluation*. What basic functionality needs to be provided by an MDQM system and which system fulfills best the imposed business requirements?
- *System landscape analysis and design*. Which MDQM systems are currently providing the required functionality, and which ones should do so in the future?
- *System landscape transformation*. What is a reasonable approach and schedule towards the future MDQM system landscape?

- *Communication*. What terminology is to be used for company-internal and external communication (e.g. with MDQM system vendors, system integrators, other companies)?

Existing literature hardly delivers any answers to these questions. While there are studies on both MDM (Smith and McKeen 2008; Loshin 2008) and DQM (Wang 1998; Batini and Scannapieco 2006; Lee et al. 2006), an integrated analysis cannot be found. Sources addressing practitioners are numerous with regard to the general significance of master data quality (Radcliffe and White 2009; White et al. 2006), recommendations for doing successful MDM (Dreibelbis et al. 2008; Loshin 2008), and comparison of system vendors (Karel 2006; Kokemüller 2009). General statements, however, on the functionality of MDQM systems are missing.

1.2 Research question and contribution

Against this background, the article puts up the following research question: What functionality is to be provided by a system supposed to improve and maintain the quality of master data? To answer this question, the paper follows the principles of Design Science Research (DSR) (Hevner et al. 2004; Peffers et al. 2008) in order to develop and evaluate a functional reference model for MDQM.

In general, a reference model is an information model that can be used not just in one specific context but in various contexts (Becker et al. 2004; Fettke and Loos 2007). For a specific class of companies, a reference model claims to be generally applicable and to serve as a predefined pattern to cope with practical problems (Rosemann and Schütte 1999; Schütte 1998). The design of reference models has a long tradition in the Information Systems (IS) research community, particularly in Europe (Frank et al. 2008, p. 391).

From an epistemological perspective, the MDQM reference model is an artifact and, thus, the result of design oriented research (March and Smith 1995; Nunamaker et al. 1991). DSR aims at designing artifacts according to scientific principles, in order to be able to solve practical problems (Hevner et al. 2004; March and Smith 1995). Hence, the reference model contributes both to the scientific and the practical state of the art:

- *Contribution to practical state of the art*. The reference model supports companies trying to overcome the challenges listed in the previous section. It helps create a common terminology both for internal and external communication. Furthermore, it offers an instrument for evaluating existing and identifying required MDQM functionality, and it allows comparing of different commercial software systems on a vendor-independent basis.
- *Contribution to scientific state of the art*. The description of the design process and of concrete design decisions allows scientific validation of the artifact presented as well as its extension by aspects previously not sufficiently considered or differentiated. Furthermore, the reference model represents an abstraction of an information system in the field of MDQM. Hence, it forms a "theory for designing" (Gregor 2006).

The remainder of the paper starts with an introduction to DQM and MDM as scientific areas of interest, followed by a brief discussion on related studies. After that, the paper outlines the design process, introduces the reference model, and explains concrete design decisions with regard to one specific function group, namely *Cross Functions*. For the purpose of demonstration the paper then describes how the reference model was applied at consumer goods manufacturer Beiersdorf in the context of a participative case study (Baskerville 1997). After that, the results of the case study and the design process are evaluated against a number of dimensions for evaluation (Frank 2007). The paper concludes with a short summary and recommendations for further research on the topic.

2 Related work

2.1 Data quality management

Data quality has been the subject of a number of research studies, some of which come up with lists and categories of data quality dimensions as the result of empirical research (Wang and Strong 1996), while others present practitioners' experiences (English 1999; Loshin 2001; Redman 1996) or draw theoretical conclusions (Price and Shanks 2005; Wand and Wang 1996). Despite all differences, there is consensus in the specification of the term *data quality* as *the data's fitness for use*, i.e. evaluating the quality of data depends on the context the data is used in and the user the data is used by.

Data quality management comprises initiatives for improving the quality of data (Batini and Scannapieco 2006). In doing so, DQM goes beyond mere reactive initiatives (i.e. identification and fixing of data defects) (Shankaranarayanan and Cai 2006). Rather, DQM aims at carrying out preventive initiatives in order to be able to ensure data quality by means of an iterative process including steps to define, measure, analyze and improve data quality as well as to design appropriate framework conditions for DQM (English 1999; Wang 1998; Eppler and Helfert 2004). An overview of various DQM approaches is given by Batini et al. (2009).

2.2 Master data management

Master data specifies the essential business entities a company's business activities are based on. Such entities are, for example, business partners (customers, suppliers), products, or employees (Smith and McKeen 2008). Basically, master data can be differentiated by three concepts: *master data class*, *master data attribute,* and *master data object* (Loshin 2008). A master data object represents a concrete business object (an automobile manufactured in a certain plant at a certain point in time, for example), and it specifies selected characteristics of this business object (color, features, or price, for example) by means of attributes. Attributes selected for representation of a specific class of business objects (customers or products, for example) constitute a master data class (which usually is specified by a data model). From the perspective of data modeling, a master data object is an

instance of a master data class, which is created by assigning data values (a sequence of figures, for example) to a master data attribute (to the attribute *phone number* of the data class representing customers, for example).

Master data management comprises all activities for creating, modifying or deleting a master data class, a master data attribute, or a master data object (Smith and McKeen 2008; White et al. 2006), i.e. the modeling, provision, quality management, maintenance, and archiving of master data. All these activities aim at providing master data of good quality (i.e. master data that is complete, accurate, timely, and well-structured) for being used in business processes (Loshin 2008; Karel 2006).

2.3 MDM and DQM

The fact that data quality has been specified as one objective of MDM has led to DQM being considered in a number of studies as only one of several functions of MDM (DAMA 2009; Dreibelbis et al. 2008). Taking this view, analysis of DQM application systems is limited to reactive DQM only (White and Radcliffe 2008). Looking at both areas in a differentiated manner, however, it can be seen that there are activities in both areas mutually affecting each other, so that any subordination of one area to the other is inappropriate. Particularly preventive DQM activities [data governance, for example (Weber et al. 2009; Khatri and Brown 2010) or business metadata management (Burnett et al. 1999; Marco 2000)] help structure a company's MDM as they specify master data or define responsibilities for data maintenance. Hence, preventive DQM does have an effect on the design of a company's MDM. Taking this into consideration the reference model presented in this paper does not aim at ranking the two concepts of MDM and DQM against each other. Rather, it specifies business requirements on the functionality of an application system supporting both DQM and MDM activities in order to be able to provide master data of good quality for being used in business processes.

3 Research approach

3.1 Overview

A reference model is defined as the result of a design process performed by a modeler who specifies (at a certain point in time) general elements of a system so that it serves as a reference point in the design of an information system. A reference model represents a class of use cases and can be used for the development of company-specific models (Schütte 1998, pp. 69–74). Reference models can be distinguished by application domain (e.g. retail industry, MDQM), modeling language (e.g. Entity-Relationship Model, Unified Modeling Language), size, design process, and evaluation strategy (Fettke and Loos 2004).

The paper at hand describes the model design process and the resulting reference model for MDQM functionality. According to the ARIS concept [org. *Architektur Integrierter Informationssysteme* (German), Architecture of Integrated Information

Systems], which distinguishes between four *descriptive views* (functional, data, control, and organizational view) and three *descriptive layers* (business design, technical design, and implementation layer) (Scheer 1992, 1997; Scheer et al. 2005), the functional reference model for MDQM represents the business design of a functional view on MDQM systems. It is therefore purpose-oriented and task-oriented (rather than data-oriented, for example), while it disregards technical and implementation aspects.

The work presented in the paper follows in general the guidelines for DSR proposed by Hevner et al. (2004). In particular, the design process is based on the principles of the Design Science Research Methodology (DSRM) (Peffers et al. 2008), suggesting a sequential design process comprising iterations of design and evaluation cycles (Simon 1998; Hevner et al. 2004) as well as various points of entry. The model design process followed in this research represents a problem-centered initiation of the DSRM process model, as the need for doing research on the topic was identified by focus groups (see Sect. 3.2). The need for doing research on the topic was neither announced in advance, nor did it result from reported shortcomings of an existing artifact.

The research context is formed by the Competence Center Corporate Data Quality (CC CDQ) at the Institute of Information Management at the University of St. Gallen.[1] Since 2006, researchers of the Institute of Information Management, together with a number of partner companies, have been developing design artifacts in the field of corporate data quality management.

3.2 Research process

As proposed by the DSRM process model, the design of the reference model was carried out in six steps.

The *first step*, which was carried out between January and December 2008, aimed at identifying the problem and motivating the research. As outlined in Sect. 1.1, the research described in this paper was mainly motivated by the identification of a number of challenges in the practitioners' community. In 2008 the market for MDQM was characterized by major consolidation activities. IBM, for example, acquired Ascential (IBM 2005), and SAP bought Business Objects (SAP 2007b). In parallel to this ongoing market consolidation, practitioners who participated in the CC CDQ continuously articulated the demand for support with regard to the four challenges mentioned above. Managerial publications supported the observed results. Business analyst company Gartner, for example, documented "common questions on data integration and data quality" from their 2008 MDM summit. Exemplary questions were (Friedman 2009):

- "What is the relative strength of MDM solution providers with regard to data quality functionality? Are they as competent as the leading best-of-breed data quality tool vendors?"

---

[1] See http://cdq.iwi.unisg.ch/.

- "Has market consolidation already reached the point where the advantages of single-vendor stacks for MDM (including data integration and data quality tools) outweigh the advantages of a best-of-breed strategy?"

The *second step* in the research process was about the definition of the objectives of the solution. The objectives of the research resulted from the identification of the practical challenges mentioned above and the realization that the existing knowledge base was not able to deliver appropriate responses to these challenges (see Sect. 1). This gap demanded for a "technique of practical utility"—which is one perception of reference models according to Fettke and Loos (2004).

Moreover, the reference model has to comply with the *Guidelines for Orderly Modeling* (*GOM*), i.e. correctness, relevance, economic viability, clarity, comparability, and systematical design (Becker et al. 1995). The GOM have been used for evaluation of reference models and have been specialized for that purpose (Schütte 1998). Summarizing, the proposed solution is a reference model which on the one hand provides a solution for overcoming the challenges mentioned above and on the other hand is designed according to the GOM. The proposed solution was confirmed within a focus group interview on December 3, 2008 (focus group B, see "Appendix").

The *third step* comprised the design activities which followed the general principles of reference modeling (Becker et al. 2002; Fettke and Loos 2004; Schütte 1998). The theoretical foundations of the model's design itself are outlined below in Sect. 4.

The design process was carried out in three iterations. The first version of the reference model was built on the basis of an integrated state of the art analysis. The analysis included both a literature review and a functional analysis of selected software systems in the MDQM domain, namely from IBM, Oracle, and SAP (see Table 1). The system analysis follows a recommendation by Hevner et al. (2004) to look into existing instantiations when analyzing the state of the art. The relevant criterion for the selection of systems was that each system had to offer broad MDQM functionality, because the first design iteration aimed at identifying as many functions as possible in order not to neglect significant aspects.

The second and the third design iteration were based on the results from focus groups A (2nd version) and C and D (3rd version). All three focus groups were used for demonstration and evaluation purposes (see below). Focus groups are a useful research method when researchers want to learn about the degree of consensus on a

Table 1 Knowledge sources for 1st design iteration

Sources of knowledge	Analyzed items
Books	Dreibelbis et al. (2008), English (1999), Heilig et al. (2006)
Industry associations	DAMA (2009)
Functionality of software systems	IBM (2006), Oracle (Butler 2007; BEA 2006), SAP (2008), TIBCO (2008)
Reports from analysts, consultancies etc.	White et al. (2006), White (2008), White and Radcliffe (2008)

topic (Morgan and Krueger 1993, p. 17). Therefore, they fit well the purpose of a reference model, which aims to represent common understanding of a topic. A similar use of focus groups in the process of reference model design can be found in the work of Sinha et al. (2004) and Schelp and Winter (2007).

The *fourth step* of the design process aimed at demonstrating the applicability of the reference model. Three different demonstration activities were used:

- Focus groups A, C, and D (see "Appendix" for details);
- Participative case study at Beiersdorf (see Sect. 5 and "Appendix" for details);
- Applicability check with three software vendors, namely IBM, SAP, and TIBCO.

In total, 31 persons participated in the three focus groups. 55% of them represented user companies, 39% consultancies, and 6% software companies. This means that all relevant groups of the MDQM community were involved in the process. Moreover, the third version of the reference model was reflected against the expertise of three software vendors (IBM, SAP, TIBCO) in order not to neglect of existing instantiations and also not to include functionality which was deemed "technically infeasible" by software providers.

Apart from that, the reference model was applied in a "real life" context during the participative case study (cf. Baskerville 1997) at consumer goods manufacturer Beiersdorf. The case study was carried out between February and October 2010.

In the *fifth step* the reference model was evaluated. Activities included:

- Focus groups A, C, and D;
- Participative case study at Beiersdorf;
- Multi-perspective evaluation according to the guidelines proposed by Frank (2007) (see Sect. 6);
- Compliance with GOM.

Due to the active role the researchers took in the process of adapting the reference model in the course of the participative case study, additional evaluation activities were required. Therefore, the paper uses both the evaluation criteria for reference modeling proposed by Frank (2007) and the GOM (Schütte 1998) for a more inter-subjective artifact evaluation.

The *sixth step* included communication activities. Both Hevner et al. (2004) and Peffers et al. (2008) stipulate that DSR results must be disseminated both in the practitioners' and the scientific community. While the former is addressed by a working paper (Otto and Hüner 2009) and a managerial publication (Otto et al. 2009) (which includes a model overview only), the paper at hand aims at making the research available for the scientific body of knowledge. First, it describes the reference model itself so that it can be used, extended, and evaluated by future research. Second, the paper outlines the research process to make it verifiable and repeatable for other researchers.

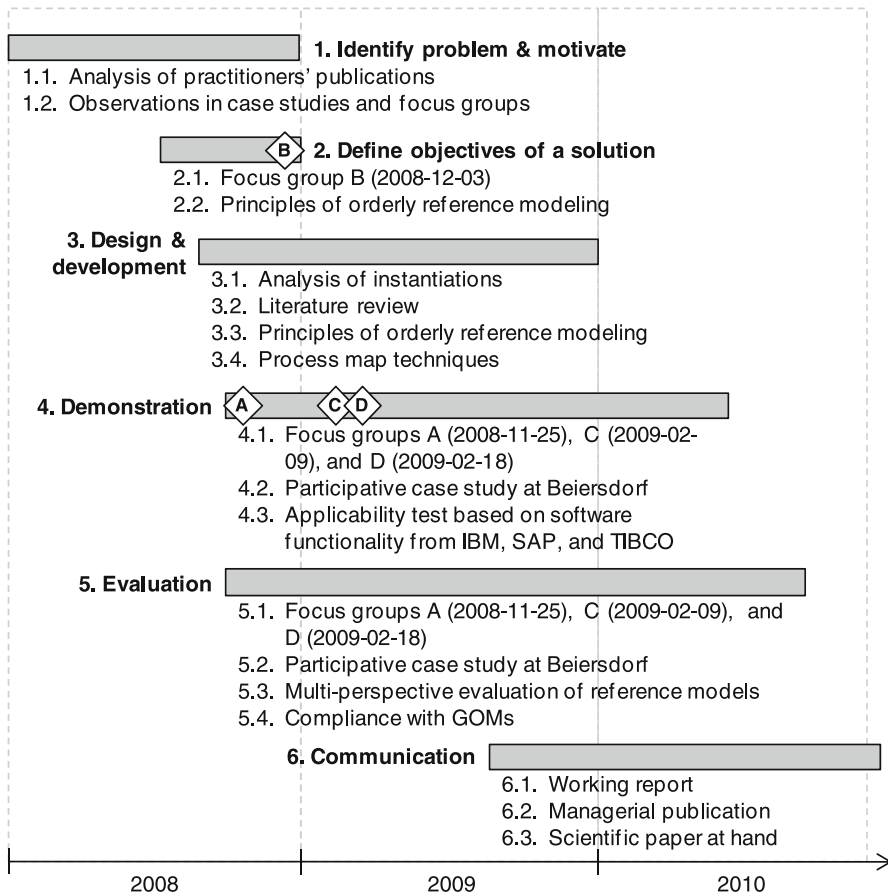Figure 1 summarizes the six steps of the research process.
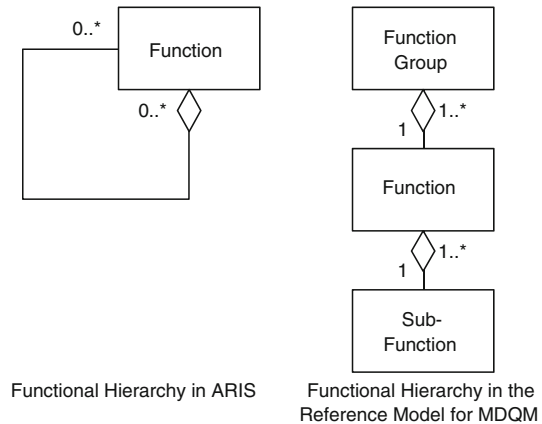
Fig. 1 Research process overview

4 Reference model design

4.1 Design foundations

The design of the functional reference model for MDQM follows the ARIS conventions for the functional view of information systems (Scheer 2001, pp. 21–38), according to which functions should be hierarchically structured. The reference model presented in this paper consists of a three-level structure formed by function groups, functions, and sub-functions (see Fig. 2).

Functions of the reference model for MDQM are grouped in function groups. A function group consists of one or more functions, whereas each function is assigned to only one function group. Functions themselves consist of sub-functions. The cardinalities between functions and sub-functions are analogous to those between function groups and functions. The use of three hierarchical levels conforms with the modeling propositions outlined in ARIS (Scheer 2001, p. 25).

Fig. 2 Modeling functional
hierarchies

Functional Hierarchy in ARIS   Functional Hierarchy in the
                                       Reference Model for MDQM

In general, there are three criteria to form functional hierarchies, namely
performance, object, and process. The reference model for MDQM uses process
criteria, grouping functions in function groups and sub-functions in functions, based
on their purpose-oriented and task-oriented relationships.

The visual representation of the reference model follows the principles of process
maps, which in general aim at identifying and representing similar processes, sub-
processes and functions (Heinrich et al. 2009) in a tabular form. This concept is also
widely used in the practitioners' community, for example in SAP's business maps.
In particular, technology-related business maps (e.g. for SAP NetWeaver) use a
tabular design (SAP 2007a). The tabular presentation format is reused in the present
paper because it was deemed appropriate to ensure high comprehensibility and
acceptance by potential users of the reference model.

4.2 Model overview

In total, the reference model comprises 6 function groups, 19 functions, and 72 sub-
functions. Figure 3 shows the function group and the function level of the model (all
72 sub-functions are explained in the "Appendix").

The following listing describes the six function groups.

- *Master Data Lifecycle Management*. A master data object's lifecycle starts with
  its creation during business operations and ends with its deactivation and/or
  archiving (Redman 1996; Loshin 2001). *Master Data Lifecycle Management*
  describes all activities data users or data managers do with master data during its
  entire lifespan (Reid et al. 2007). Self-explaining functions, such as *Create* or
  Update, are not stated explicitly.
- *Metadata Management and Master Data Modeling*. Basically, metadata
  specifies data properties and the meaning of data (of master data, for example).
  In doing so, metadata both specifies data structures and—in the form of
  unambiguous specifications—ensures correct usage of data throughout an

Master Data Lifecycle Management	Data Creation	Data Maintenance	Data Deactivation	Data Archiving
Metadata Management and Master Data Modeling	Data Modeling	Model Analysis	Metadata Management	
Data Quality Assurance	Data Analysis	Data Enrichment	Data Cleansing	
Master Data Integration	Data Import	Data Transformation	Data Export	
Cross Functions	Automation	Reports	Search	Workflow Management
Administration	Data History Management	User Management		

Fig. 3 Function groups (*left column*) and associated functions

organization (Burnett et al. 1999; Marco 2000; Tozer 1999). From an MDM perspective, metadata comprises all the information necessary for efficient management and effective usage of master data. According to the definition above, master data modeling means creation of technical metadata (data types, relationship multiplicities, for example).

- *Data Quality Assurance*. This function group comprises functions for preventive (i.e. to prevent potential, future data defects) and reactive (i.e. to repair data defects that have occurred) maintenance and improvement of the quality of master data. The three functions comprise sub-functions for identification of data defects and measuring data quality (*Data Analysis*), for improving data quality by comparison with and integration of external reference data (*Data Enrichment*) and for repair of data defects identified (*Data Cleansing*).
- *Master Data Integration*. This function group comprises functions supporting transfer (import and export) and structural transformation (e.g. consolidation of fields or tables) of master data.
- *Cross Functions*. This function group comprises functions that cannot be assigned to one of the other groups. Sub-functions under the function *Automation* do not provide additional functionality but offer support for being able to efficiently use other functions by making them machine processable.
- *Administration*. This function group comprises functions for user administration and the tracing of changes and modifications made.

Cross Functions	Automation	Reports	Search	Workflow Management

Automation	Reports	Search	Workflow Management
Automated Enrichment	Data Quality Reports	Dynamic Value Search	Bundling of Activities
Automated Export	Usage Statistics	Free Search	Graphical Workflow Modeling
Automated Import	Job Monitoring	Fuzzy Search	Create/Maintain Workflows
Cross-Function Automation	Audit Support		
Push and Pull Mechanisms			

Fig. 4 Functions and sub-functions of function group *Cross Functions*

4.3 Functions in detail

Figure 4 displays the four functions and the fifteen sub-functions of the function group named *Cross Functions*. As the definition and structural design of this function group was a frequent issue during the discussions in the focus group interviews and the application of the reference model, Table 2 outlines design decisions made in the course of several design iterations, leading to the structural design of the function group as shown in Fig. 4 (see "Appendix" for the description of the sub-functions).

5 Reference model application

5.1 Case study overview

A participative case study (Baskerville 1997) was used to demonstrate the reference model's applicability in a "real life" context. A participative approach allows for evaluation of the applicability of the model. The case study was conducted at consumer goods manufacturer Beiersdorf AG, headquartered in Hamburg, Germany. Operating in two business segments, namely Consumer (including brands as *Nivea*) and Medical (including brands like *Hansaplast*), Beiersdorf is a global company, generates an annual revenue of approx. EUR 5.7 billion in 150 subsidiaries, and employs about 20,000 people all over the world (Beiersdorf 2009).

The master data class investigated in the participative case study is product data, e.g. identifiers, logistics data, and product hierarchies (Saaksvuori and Immonen 2008). For the management of global product data Beiersdorf is using a central product lifecycle management (PLM) system, which was implemented in 2004. The
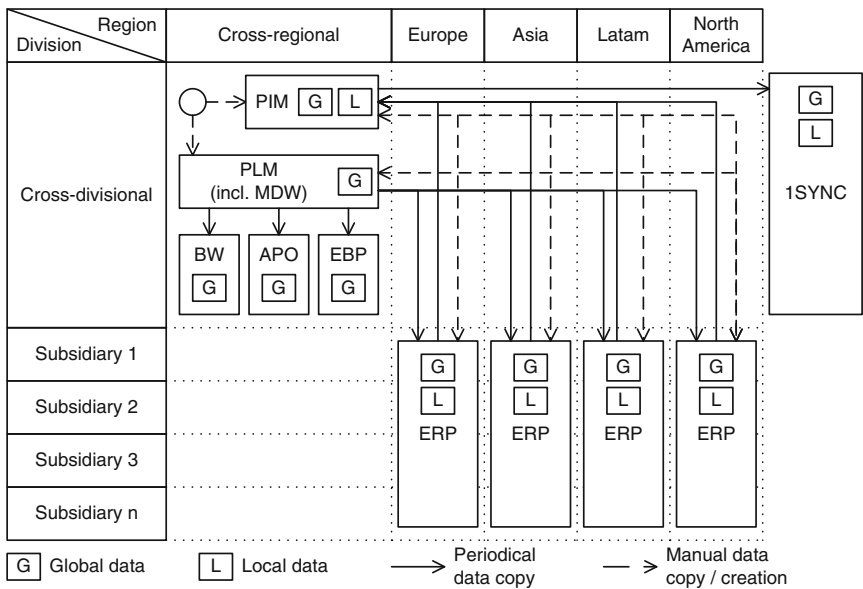
Table 2 Examples of design decisions and justifications regarding the function group *Cross Functions* (sorted chronologically)

Design decision	Justification
No explication of cross functions in the initial version of the reference model	Simplicity (principle "as simple as possible")
	Analysis of practitioners' publications (see Table 1, Fig. 1, process step 1.1)
	Observations in case studies and focus group discussions (see Fig. 1, 1.2).
Add functions for analysis and reporting (e.g. Search, Profiling, Reporting)	Analysis of practitioners' publications (Alur et al. 2007; Batini and Scannapieco 2006; Dreibelbis et al. 2008; English 1999)
	Verified by reverse engineering analysis [see Fig. 1, 4.3, (Otto and Hüner 2009)]
Add automation functions (e.g. Automated Export/Import, Automated Enrichment)	Analysis of instantiations (see Fig. 1, 1.2)
	Verified by reverse engineering analysis [see Fig. 1, 4.3, (Otto and Hüner 2009)]
Explication of cross functions (i.e. functions that support more than one function from different function groups)	Focus group A (see Fig. 1, 4.1) suggested to group cross functions (i.e. Automation, Reports, Search) analogous to Corporate Services of SAP's ERP Solution Map (SAP 2011)
	Allows differentiation of end user functionality and supporting functionality (cross functions)
	Verified by focus groups C and D (see Fig. 1, 5.1).
	Verified by reverse engineering analysis [see Fig. 1, 4.3, (Otto and Hüner 2009)]
Assign Data Quality Reports to Cross Functions (previously under Data Quality Assurance), with Graphical Analysis remaining under Data Analysis	Focus group A decided to leave sub-function Graphical Analysis under Data Quality Assurance due to its close relation with Profiling (for example, when searching for duplicates, identification of potential duplicates is the result of profiling and supports manual analysis in the process of duplicate recognition)
	Verified by focus groups C and D (see Fig. 1, 5.1)
Add Workflow Management to Cross Functions	Focus group B discussed organizational challenges in the context of MDQM and raised the requirement of supporting workflows for collaborative data creation and maintenance
	Verified by focus groups C and D (see Fig. 1, 5.1)
	Analysis of instantiations (see Fig. 1, 1.2) did not support this decision (regarding MDQM offerings), but each application system provider offers additional components for workflow management, and focus group members confirmed that these components were required for MDQM
Ignore function Automation for as-is functionality ratings by business users (see Sect. 1, use case System landscape analysis and design)	Business experts (see below, model application at Beiersdorf) had difficulties in distinguishing between some Automation sub-functions (i.e. Automated Import) and automated sub-functions (e.g. Delta Import) from other function groups

PLM system at regular intervals (i.e. every 3 h) provides new or modified product data to five regional enterprise resource planning (ERP) systems and a number of other global information systems [e.g. a decision support system (BW), a planning system (APO), and a procurement system (EBP)]. As the data is directly committed into the receiving systems, a consistent database is always ensured (Schemm 2008). Figure 5 illustrates the system landscape used at Beiersdorf for the management of product data. The system landscape depicted is typical for a global company, comprising both global systems supporting processes that affect several organizational units and local systems supporting processes within discrete organizational units (Lehmann 2003). Figure 5 gives examples of systems the functionality of which is supposed to be represented by the MDQM reference model.

Being part of the PLM system, the Master Data Workbench (MDW) provides functionality for master data creation, thereby ensuring that master data is captured by the PLM system right at the moment it is being created. The users of the system (about 150) work with input masks specifically designed to match with the product innovation process and the product data maintenance process. The whole process of master data gathering and creation is done by means of one single system. Fast and accurate data processing is ensured, since there is no media breakage and a number of PLM functions (e.g. allocation of unique identifiers, check routines) can already be used in the process of data creation.

Product data distribution to wholesalers and retailers is controlled by means of a central product information management (PIM) system. The PIM system is

1SYNC: Product data pool, APO: Planning system, BW: Business warehouse, EBP: Procurement system, ERP: Enterprise resource planning system, MDW: Master data workbench (master data creation), PIM: Product information management system, PLM: Product lifecycle management system

Fig. 5 Application landscape for managing master data at Beiersdorf (Hüner et al. 2011)

provided with global and local master data from both the PLM system and local ERP systems. The PIM system also controls the data transfer to 1Sync, Beiersdorf's data pool.

5.2 Case study approach and results

At Beiersdorf, planning, maintenance, and implementation of the system landscape is done by a subsidiary of the company, named Beiersdorf Shared Services (BSS). The majority of systems (PLM, ERP, BW, or APO systems) have been implemented with products from SAP. One consequence of this single-source strategy is that complexity regarding interfaces and license contracts could be reduced. Another consequence, however, has been a growing dependence on this particular vendor. The consolidation process in the market for DQM and MDM systems in general, together with changes in SAP's portfolio in particular, has raised a number of questions for BSS:

1. Which of the functions offered by the MDQM application landscape are required by customers of BSS (i.e. the business users at Beiersdorf)?
2. Does BSS offer all functions required by business users in the desired quality?
3. Are software systems from Beiersdorf's standard software providers available for functions that are currently provided by self-developed or modified standard software systems?
4. Are alternative products available for functions BSS is currently providing not at all or not in the desired quality?

These questions relate to the challenges introduced in the first section of this paper, namely *System evaluation* (3 and 4), *System landscape analysis and design* (1 and 2), and *System landscape transformation* (3 and 4). Thus, the case of Beiersdorf can be taken for evaluation of the reference model (cf. Frank 2007; vom Brocke 2007). The application of the reference model comprises the following four activities (with Activities 1–3 already being completed):

1. *Preparation*. The reference model was not configured. For evaluation of the 72 sub-functions (see Activity 2), a guideline explaining and illustrating each sub-function, an assessment method (see below), and a template for documentation was developed.
2. *Assessment*. In a workshop comprising one expert from BSS and three business users, each of the 72 sub-functions was rated (see below) in terms of (1) to what degree it is demanded by users, and (2) to what degree users are currently satisfied with it.
3. *As-is analysis*. For each of the 72 sub-functions the component of the system landscape providing the function was identified (PLM system, for example). Furthermore, each component was documented with regard to being a standard component, an adapted standard component, or a self-developed component.
4. *To-be concept*. The analysis of the rating results (Activity 2) and the as-is analysis (Activity 3) lead to the identification of functions that are demanded but are provided not at all or not in the quality desired. For each of these

functions it has to be stated whether it can be implemented by means of components already in place or whether a new product needs to be integrated.

The assessment method (Activity 2) helps identify functions that are highly demanded but are provided not at all or not in the quality desired. For this purpose, two metrics were defined: *demand* and *satisfaction*. Both metrics were rated by the workshop participants for each sub-function of the reference model on a Likert scale ranging from 0 to 5. For aggregation of both ratings, a coefficient matrix C (see below) was used to map the ratings on a scalar value $n = d^T \cdot C \cdot s$, which can be interpreted as a *need for action*. In this formula, d(*demand*) and s(*satisfaction*) are unit vectors with 1 indicating the respective rating [e.g. $(0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0)^T$ represents a rating of 4].

$$C = \begin{pmatrix} 0/25 & 0/25 & 0/25 & 0/25 & 0/25 & 0/25 \\ 5/25 & 4/25 & 3/25 & 2/25 & 1/25 & 0/25 \\ 10/25 & 8/25 & 6/25 & 4/25 & 2/25 & 0/25 \\ 15/25 & 12/25 & 9/25 & 6/25 & 3/25 & 0/25 \\ 20/25 & 16/25 & 12/25 & 8/25 & 4/25 & 0/25 \\ 25/25 & 20/25 & 15/25 & 10/25 & 5/25 & 0/25 \end{pmatrix}$$

An example: A function rated '4' for *demand* and '2' for *satisfaction* is given the value 12/25. The meaning of the metrics and scales was explained to the workshop participants, so that approximately equidistant rating of functions could be ensured, allowing interpretation of the level of measurement according to a ratio scale.

Figure 6 gives an overview of the ratings of the 72 sub-functions (with dark coloring indicating a high need for action) and shows the results of the as-is analysis (smaller part on the right-hand side of each box representing a function).

The as-is analysis (see Fig. 6) shows the highest need for action in *Archiving* and *Usage statistics*. Interestingly, standard software is already available to support archiving functionality, but it is obviously not used. In contrast, for usage statistics no support at all is available at present. Apart from that, Beiersdorf identified a number of functions with relatively high need for action and no software support. An example of that is *Glossary/Dictionary*, which is used to establish shared definitions for master data in the organization and to provide business context information.

Besides the identification of MDQM functions and sub-functions to be further developed at Beiersdorf, the application of the reference model has motivated a modification of the reference model itself. Workshop participants had difficulties in distinguishing between some cross functions (i.e. *Automation*) and sub-functions from other function groups (e.g. *Delta Import* and *Export*, *Profiling*). From a business perspective, it is hardly possible to decide whether a certain data record is in place (and of good quality) because of an automated import and cleansing routine or because of a manual activity performed by IT staff (BSS in the case of Beiersdorf). Therefore, sub-functions from function *Automation* (of function group *Cross Functions*) will no longer be considered for as-is functionality ratings by business users (see Sect. 1, use case *System landscape analysis and design*). However, in

Master Data Lifecycle Management	Data Creation	Data Maintenance	Data Deactivation	Data Archiving
	Conditional Entries	Check-out	Bulk Editing	Archiving
	Bulk Editing	Bulk Editing		History Control
	Plausibility Check	Plausibility Check		

High demand and low satisfaction

Metadata Management and Master Data Modeling	Data Modeling	Model Analysis	Metadata Management	
	Data Model Editing	Dependency Analysis	Business Rules Documentation	
	Graphical Modeling	Data Type Recognition	Glossary/Dictionary	
	Classification	Primary and Secondary Key Recognition	Metadata Import	
	Support of Business Standards	Relationship Recognition	Mandatory Fields Administration	
	Data Model Version Control		Metadata Publication	
			Metadata Transport	
			Metadata Visualization	

No demand or high satisfaction

Data Quality Assurance	Data Analysis	Data Enrichment	Data Cleansing
	Compliance Verification	External Reference Data	Delta Import
	Graphical Analysis	Classification Schemes	Duplicate Recognition
	Plausibility Lists	Measuring Units	Pattern Recognition
	Profiling	Multilingual Capability	Plausibility Check
		Management of Unstructured Data	Spelling Check

Master Data Integration	Data Import	Data Transformation	Data Export
	Delta Import	Field Split	Search Based Data Selection
	Import Formats	Field Merge	Delta Export
	Connectors	Data Type Conversion	Export Formats
	Virtual Integration	Pivot Tables	Connectors
			Limitation
			Preview

Cross Functions	Automation	Reports	Search	Workflow Management
	Automated Enrichment	Data Quality Reports	Dynamic Value Search	Bundling of Activities
	Automated Export	Usage Statistics	Free Search	Graphical Workflow Modeling
	Automated Import	Job Monitoring	Fuzzy Search	Create/Maintain Workflows
	Cross-Function Automation	Audit Support		
	Push and Pull Mechanisms			

Administration	Data History Management	User Management
	Data Lineage	User Interface Design
	Last User	Roles and Rights

Standard software
Custom solution
Not supported

Fig. 6 Reference model applied for rating of functions

responding to other challenges (e.g. *System evaluation*) the differentiation between function *Automation* and automated sub-functions has proven to be reasonable (see reverse engineering analysis of MDQM application systems (Fig. 1, process step 3.1). The differentiation has also been confirmed during the validation by vendors (see Fig. 1, 4.3).

6 Evaluation

6.1 Multi-perspective artifact evaluation

With a particular focus on the evaluation of reference models, Frank (2007) has proposed a framework comprising four perspectives of evaluation. This framework is used for evaluation of the reference model presented in this paper, taking into consideration the challenges described in the first section.

- *Economic Perspective*. Due to the simple structure of the reference model (three levels) and clearly defined objectives, the costs for training, adaptation and application (see *Deployment Perspective* and *Engineering Perspective*) are low (1 day for preparation and 2 days for application in the case presented). Tools supporting the processes of rating and analysis (see Fig. 6) can be created at low effort (Microsoft Excel based templates for documentation in the case of Beiersdorf, for example). Using the model does not lead to direct cost savings. However, the as-is analysis might identify unnecessary application system licenses or potentials for consolidation. The analysis of the ratings as well as comparison with other companies on the basis of the reference model might lead to better support of required functionality. Both the focus group interviews and the final validation by application system providers have shown that the reference model is capable of substantially simplifying exchange of knowledge and comparison of products.
- *Deployment Perspective*. The focus group interviews, the validation of application system providers, and the application of the reference model at Beiersdorf have shown that the model is easy to understand and well applicable. Any rejection of the model due to the fact that it was developed externally (the *not-invented-here-syndrome*) could not be observed.
- *Engineering Perspective*. Applying the reference model at Beiersdorf has shown that the model can be used to overcome the challenges described. The model's simple structure (see *Deployment Perspective*) ensures its easy adaptability (cf. vom Brocke 2007). After the model had been applied at Beiersdorf, two functions (*Data Sourcing and Data Delivery*) for cross-company exchange of master data (product data for external production of artworks and electronic delivery of them, for example) were added under the function group *Master Data Integration* in order to be able to take into account company specific requirements of Beiersdorf (administration of subscriptions at the interfaces of different artwork agencies, for example) for future ratings.

- *Epistemological Perspective*. The validation by application system providers (see Fig. 1, 4.3.) as well as the application of the reference model at Beiersdorf (see Fig. 1, 5.2.) has shown that the model is capable of abstracting and representing reality (i.e. MDQM systems). Critical distance is ensured by explication of use cases. By taking into account only one of four ARIS views and one of three ARIS levels, the development and implementation of an application system, for example, is no adequate use case. Moreover, explication of the model design process ensures that scientific principles are followed (such as verifiability and reproducibility of the artifact).

6.2 Compliance with GOM

Guidelines for orderly modeling can be used to evaluate the validity of artifact design. They allow for assessing whether the reference model complies with the specifications developed in the second step of the design process (see above) (Table 3).

Table 3 GOM compliance

GOM	Description	Assessment
Correctness	Model consistency, completeness	The three focus groups did not identify any missing components. However, the design of the cross functions remains questionable
	Semantic correctness	
	Naming conventions	The reference model follows conventions of the ARIS functional view
Relevance	Appropriateness of model boundaries	The focus on MDQM is relevant, as the motivation in Sect. 1.1 of the paper shows. The level of abstraction has proven appropriate both in the focus groups and the participative case study
	Appropriate level of abstraction	
Economic viability	Reasonable modeling effort	The model has been developed without the need for significant expenditures. Its application fulfils the criteria of the economic perspective according to Frank (2007) (see above)
Clarity	Comprehensibility, readability	Becker et al. (1995, p. 438) concede the limited verifiability of this guideline. The reference model, however, was understood well by subject matter experts in the focus groups, by vendor representatives, and by case study participants
Comparability	Compatibility to other models	The compatibility to other models has not been evaluated
	Support of "as is" and "to be" models	The support of comparing "as is" and "to be" situations in a company has been demonstrated in the case study at Beiersdorf
Systematical design	Decomposition into views, layers etc.	The reference model has a hierarchical structure consisting of three levels. This design has been used in similar cases, e.g. in SAP's business maps (SAP 2007a)

7 Conclusion and outlook

The paper describes a functional reference model for MDQM. The model design process spanned the six steps as proposed by DSRM and includes several design and evaluation cycles.

The reference model is beneficial with regard to both the advancement of the scientific state of the art and the state of the art in practice. Practitioners can use the reference model as an instrument supporting the analysis and design of their MDQM system landscapes. Furthermore, the model facilitates evaluation of software products, and it facilitates both inner-company and cross-company communication. From a researcher's perspective, the reference model is a representation of an information system, as it specifies business user requirements for MDQM system functionality, and as it delivers new knowledge in terms of representing reality (cf. Frank 2007). Explication of the research process allows verification, correction, and differentiation of this representation.

The reference model has its limitations due to its focus on the business layer of the functional view of MDQM (Scheer 1992, 1997) and due to the fact that other ARIS views and levels were not modeled. Hence, the application of the reference model is restricted to use cases similar to the one described. Besides scientific validation of the reference model, further research on the topic should aim at extending the model and adding to it more views and levels. The authors of this paper think that especially the control view and the organizational view offer potential for designing relevant artifacts. Based on case studies, generic character-istics of DQM and MDM organizations (roles and responsibilities, for example) could then be identified to constitute the basis for conceptualizing rights and roles as the reference model's organizational view. Besides, interdependencies between individual functions of the reference model could be identified and typical MDQM activities (adapting the master data architecture, identifying and clearing duplicates, measuring data quality, for example) could be described.

Appendix

Focus group details

Table 4 lists the participants of a focus group interview of the MDM working group of the German-speaking SAP User Group, taking place on November 25, 2008, in St. Leon-Rot, Germany. The focus group interview started at 9:45 and ended at 12:15. After an introduction of 45 min about the context, motivation and current work progress, the first version of the reference model was presented to the 17 focus group participants. The presentation comprised a model overview analogous to

Table 4 Participants of focus group A

Company	Type	Participant's function in the organization
ABeam consulting (Europe) B.V.	Consulting	Senior Manager
Adolf Würth GmbH & Co. KG	User	Head of Order Processing and Processes
aseaco AG	Consulting	Managing Consultant
cbs Corporate Business Solutions Unternehmensberatung GmbH	Consulting	Senior Project Manager Corporate Data Management
Deloitte consulting GmbH	Consulting	Senior Manager
gicom GmbH	Consulting	Managing Director
IBSolution GmbH	Consulting	Business Development Manager SOA, Director SOA
IMG AG	Consulting	Consulting Manager
ISO Software Systeme GmbH	Software vendor	Head of Division DQM Sales SAP
Karl Storz GmbH & Co. KG	User	Head of International Master Data Management
Koch BNC AG	Consulting	Consulting Manager
Mieschke Hofmann & Partner	Consulting	Head of Competence Center Strategy, Architecture & Methods
Roche diagnostics GmbH	User	Head of Global Material Master Management
SAP Deutschland AG & Co. KG	Software vendor	MDM Consultant
Steria Mummert consulting AG	Consulting	Principal Consultant
T-Systems Enterprise Services GmbH	User	Project Manager and Consultant
Voith Paper Holding GmbH & Co. KG	User	Business Processes & Information Technology, Head of Master Data Management

Fig. 3 as well as a detailed view of all six function groups. The participants were asked for their feedback with regard to completeness, consistency, usefulness, and correctness of the model. The focus group interview was recorded and transcribed.

Table 4 shows the participants of focus group B, which was part of a regular workshop of the CC CDQ. The focus group took place on December 3, 2008, in Munich, Germany, and involved 12 participants. The focus group was not specifically set up on the topic of MDQM functionalities, but was aiming at short-term and medium-term research activities in the field of corporate data management. The participants were asked to name and explain three short-term (for the first 18 months) and one mid-term (18–24 months) research priorities. After that a moderated discussion identified seven focus topics, of which one was the Functional Reference Model for MDQM.

Table 5 shows the participants of focus group C, which was part of an "in-depth workshop" on corporate data quality management. The focus group took place from 15:45 to 16:45 on February 9, 2009, in Cologne, Germany. The focus group was set up and conducted analogous to focus group A. The results were documented on a flip-chart.

Table 5 Participants of focus group B

Company	Type	Participant's function in the organization
ALSTOM Power	User	Chief Data Officer
B. Braun Melsungen AG	User	Head of Central Material Master Agency
BMW AG	User	Manager IT System Landscape
DB Netz AG	User	2 Team members Infrastructure Data Management
Deutsche Post Adress GmbH & Co. KG	User	Sales Manager
Deutsche Telekom AG	User	Team member Master Data Management
EFQM	Standards organization	Chief Architect
IBM Deutschland GmbH	Software vendor	Manager Enterprise Architecture Management Solutions
Nestlé S.A.	User	Manager Business Data Excellence
Raiffeisen Schweiz	User	Manager and team member Data Management
Syngenta Crop Protection AG	User	Team member Master Data Management
Tchibo GmbH	User	Manager Master Data Management

Table 6 Participants of focus group C

Company	Type	Participant's function in the organization
Client Vela GmbH	Consulting	Partner
Corning Cable Systems GmbH & Co. KG	User	Global Data Management Organization Leader
Helsana Versicherungen AG	User	Head of Quality Management PK, Solution Designer (technical/business) DWH
Just.dot GmbH	Consulting	Managing Director
Mars Service GmbH	User	Business Data Manager Europe, Supply Chain Development

Table 6 shows the participants of focus group D, which was part of a regular workshop of the CC CDQ which formed the context of this work. It took place between 10:30 and 11:30 on February 18, 2009, in St. Gallen, Switzerland. The focus group was set up and conducted analogous to focus groups A and C. The results were documented on flip-charts (Table 7).

Case study details

The case study at Beiersdorf uses interviews for data collection. Table 6 gives details on the six interviews conducted (Table 8).

Table 7 Participants of Focus group D

Company	Type	Participant's function in the organization
ABB Information Systems Ltd.	User	Master Data Consultant and Headquarter IS Architect
Aesculap AG	User	SAP Inhouse Consultant Development and Master Data Management
B. Braun Melsungen AG	User	Head of Central Material Master Agency
Bayer CropScience AG	User	Integration Manager Enterprise Master Data Management
Beiersdorf AG	User	Head of Supply Chain Data Process Management
Beiersdorf Shared Services GmbH	User	Head of Team BSS Master Data
DB Netz AG	User	Head of Strategic Infrastructure Data Management, Railway Geographical Data, Strategic Infrastructure Data Management, Railway Geographical Data
Geberit International AG	User	Head of Group Product Data Management
SBB Cargo AG	User	Head of Business Process Management

Table 8 Case study interviews

Date	Organizational unit	Meeting type (and topic)	Number of participants	Duration (h)
23 Feb 2010	BSS	Project meeting (conference call, project planning)	1 (BSS)	1
30 Mar 2010	BSS, SC DPM	Assessment workshop (onsite)	2 (BSS), 3 (SC DPM)	4
7 Apr 2010	BSS, SC DPM	Assessment workshop (onsite)	2 (BSS), 3 (SC DPM)	4
31 May 2010	BSS	Project meeting (conference call, discussion assessment analysis)	1 (BSS)	2
1 Sep 2010	BSS	Project meeting (conference call, knowledge sharing planning)	2 (BSS)	1
21 Oct to 8 Nov 2010	BSS, 7 manufacturing and consumer goods companies	Survey (sent by email)	1 (BSS), representatives of 7 peer group companies	NA

Functional details of the reference model

Table 9 describes all 72 sub-functions of the reference model.

Table 9 Complete reference model overview

Functions	Description
Master data lifecycle management	
Data creation	
Conditional entries	Allows modeling of relations between master data classes that change depending on values of the associated classes. Example: Discount rates for different purchase quantities
Bulk editing	Allows to create a number of similar data objects (or particular attributes) at a single time, i.e. the creation process does not have to be executed for single data objects individually
Plausibility check	Ensures that no invalid data are entered in enterprise systems
	Could use reference lists that contain correct addresses, correct names etc.
Data maintenance	
Check-out	Prevents data objects from being edited by other users
Bulk editing	Allows to edit a number of data objects (or particular attributes) at a single time, i.e. an editing process does not have to be executed for single data objects individually
Plausibility check	Ensures that no invalid data are entered in enterprise systems
	Could use reference lists that contain correct addresses, correct names etc.
Data deactivation	
Bulk editing	Allows to delete/deactivate a number of data objects at a single time, i.e. an editing process does not have to be executed for single data objects individually
Data archiving	
Archiving	Allows to persistently store data objects for a defined period of time
	Supports compliance with relevant legal provisions.
History control	Allows to archive different versions of any piece of master data
	Ensures that any data object can be reconstructed as it was at a certain point in time
Metadata management and master data modeling	
Data modeling	
Data model editing	Allows to modify and adapt classes of master data, in order to, for example, add new attributes or mark existing attributes as mandatory fields
Graphical modeling	Data models can be created using graphical symbols (e.g. connecting types of data by lines indicating a relation between these types)
Classification	Allows to group and categorize master data
	Assigning data objects to certain categories must not necessarily be unambiguous
Support of business standards	Allows to implement business standards or to take advantage of options offering integration (e.g. import of an XML based standard as a data class for customer data)

Table 9 continued

Functions	Description
Data model version control	Allows to archive different versions of any data model
	Ensures that a data model can be reconstructed as it was at a certain point in time
Model analysis	
Dependency analysis	Verifies the effect of a change in the data structure of a certain data class (e.g. deletion of an attribute) on other data classes
Data type recognition	Allows automatic recognition of data types of existing data objects and automated recognition of existing data models (e.g. when consolidating two customer data sets)
Primary and secondary key recognition	Allows automatic identification of single attributes suited to work as the primary key
	Checks the key integrity (e.g. the unambiguousness of a foreign key attribute)
Relationship recognition	Relations between types of data are automatically recognized
	Supports consolidation of different data inventories
Metadata management	
Business rules documentation	Supports the communication of business rules (may refer to strategic decisions or to system activities), in order to simplify their use and to keep their definitions up to date
Glossary/dictionary	Clearly defines a company's central business objects or other elements needed to ensure smooth business processes (e.g. SAP fields)
Metadata import	Allows to consolidate metadata of different formats. Important as metadata are often are stored in distributed systems and in heterogeneous, partially unstructured formats
Metadata transport	Allows automatic transfer of metadata from test systems to transaction systems. Important as data structures usually are created in test systems.
Mandatory fields administration	Allows central configuration and administration of input fields specified as mandatory fields (e.g. for SAP transactions)
Metadata publication	Metadata (e.g. business rules or information about business objects) are made available for being used in enterprise systems, where they can be requested with minimum effort
Metadata visualization	Uses metadata (threshold values, business rules templates etc.) in order to graphically display complex phenomena in a simplified manner (e.g. diagrams, scorecards)
Data quality assurance	
Data analysis	
Compliance verification	Allows to verify master data against certain guidelines or legal provisions
Graphical analysis	Allows graphical representation of profiles created by means of Profiling (e.g. by illustrating the frequency distribution of values of an attribute)
Plausibility lists	Provides a basis for other functions (Profiling, Plausibility Check)
	May contain reference data (e.g. company addresses) or data formats

Table 9 continued

Functions	Description
Profiling	Allows to analyze data and, based on predefined rules, to create a statistical profile regarding compliance with such rules (basis for e.g. Duplicate Record Recognition)
Data enrichment	
External reference data	Allows to substitute missing data by external data (e.g. a company address register) or to match existing data with external data in order to detect errors in one's database
Classification schemes	Supports the use of classification systems (for example eCl@ss, ETIM) for corporate master data management
Measuring units	Supports conversion of measuring units (e.g. attributes of dimension or weight)
Multilingual capability	Allows to make master data and metadata available in various languages at constant data consistency
Management of unstructured data	Allows efficient administration of unstructured data (e.g. artwork) and their relations with master data objects as well as efficient provision of such data
Data cleansing	
Delta import	Allows to import data created or modified since the previous import (the Delta)
	Identification of the Delta can be useful to search for duplicate records, for example
Duplicate recognition	Allows to search for duplicate records
	Generates warnings during data entry indicating data duplication
Pattern recognition	Identifies certain patterns in data repositories
	Patterns allow to define expected data structures or invalid entries
Plausibility check	Ensures that no invalid data are entered in enterprise systems
	Could use reference lists that contain correct addresses, correct names etc.
Spelling check	Corrects typical mistakes occurring during data entry
	The function can be supported by reference lists used also for Plausibility Check
Master data integration	
Data import	
Delta import	Allows to import data created or modified since the previous import (the Delta)
Import formats	Ensures that only such data can be processed the format of which is understood or which are converted into a known format during the importing process
Connectors	Allows to create new interfaces for, importing data of a format originally not supported (usually offered as transformation languages (e.g. XSLT) or APIs)
Virtual integration	Allows to temporarily bring together data from different source systems without needing to copy them into a common database
Data transformation	
Field split	Allows to split values of one field into several values, following predefined rules (e.g. by a separator '_', or ';')

Table 9 continued

Functions	Description
Field merge	Allows to merge values of several fields
Data type conversion	Allows to consolidate data on the basis of a standard data type (e.g. texts 256 characters long or 64 bits wide)
Pivot tables	Allows to restructure data classes structured in tables (e.g. by new ordering schemes or inserting rows and columns)
Data export	
Search based data selection	Allows the explicit selection of data objects to be exported from a list (result of a search query)
Delta export	Allows to export data created or modified since the previous export (the Delta)
Export formats	Provides the data formats supported for data export and ensures that data are transferred to transaction systems again after being processed in one way or the other
Connectors	Allows to create new interfaces for exporting data of a format originally not supported (usually offered as transformation languages (e.g. XSLT) or APIs)
Limitation	Allows to export only a certain data set, what might be helpful in the context of test, for example to estimate the result of a cleansing initiative
Preview	Allows to view data to be exported as they will be provided
Cross-function	
Automation	
Automated enrichment	Allows automation of enrichment functionality (e.g. comparing data with external reference data, converting measurement units)
Automated export	Allows (together with Automated Import) to build a system for automated exchange of master data between a test system and transaction systems
Automated import	Allows (together with Automated Export) to build a system for automated exchange of master data between a test system and transaction systems
Cross-function automation	Allows automated execution of various, linked functions in a certain sequence (e.g. workflows that do not require human involvement)
Push and pull mechanisms	Allows to apply both push-mechanisms and pull-mechanisms for automated data import and export
Reports	
Data quality reports	Allows to illustrate the results of data analyses, e.g. by diagrams to be used in dashboards, or by preconfigured templates for management reports
Usage statistics	Allows to record in real-time who is using or requesting which data
Job monitoring	Allows to monitor automated functions and assess them by various indicators (e.g. processing time, error rate)
Audit support	Helps create (e.g. by providing templates or preconfigured analyses) reports demanded by legal provisions

Table 9 continued

Functions	Description
Search	
Dynamic value search	Allows to search for and identify data objects by means of known attribute values
	Supported by dynamic sorting and filtering mechanisms
Free search	Allows to make full-text queries across the entire database. Search results are provided in a ranking list starting with the result supposed to be of highest relevance
Fuzzy search	Provides an extension of Free Search in terms of including similar words and synonyms into the search process (e.g. the name Maier/Meier, or München/Munich)
Workflow management	
Bundling of activities	Allows to bundle several activities within a single MDM workflow
Graphical workflow modeling	Allows to model workflows by means of graphical symbols
Create/maintain workflows	Allows to manage sequences of activities across processes and departments
	Along the entire data lifecycle numerous activities are executed by numerous people
Administration	
Data history management	
Data lineage	Allows to trace back the origin of pieces or sets of master data
	Important if master data from various, distributed information systems are consolidated
Last user	Allows to identify the person who did the last modification in a set or piece of master data or who used a set or piece of master data last
User management	
User interface design	Allows to adapt the graphical user interface to meet role specific requirements
Roles and rights	Allows to define roles and to assign entitlements to execute certain activities by such roles

References

Alur N, Joseph R, Mehta H, Nielsen JT, Vasconcelos D (2007) IBM WebSphere information analyzer and data quality assessment. IBM, Armonk

Baskerville RL (1997) Distinguishing action research from participative case studies. J Syst Inf Technol 1(1):24–43. doi:10.1108/13287269780000733

Batini C, Scannapieco M (2006) Data quality. Concepts, methodologies and techniques. Springer, Berlin

Batini C, Cappiello C, Francalanci C, Maurino A (2009) Methodologies for data quality assessment and improvement. ACM Comput Surv 41(3):1–52. doi:10.1145/1541880.1541883

BEA (2006) BEA AquaLogic™ data services platform. BEA Systems, Inc, San Jose

Becker J, Rosemann M, Schütte R (1995) Grundsätze ordnungsmäßiger Modellierung (Guidelines for orderly modeling). Wirtschaftsinformatik 37(5):435–445

Becker J, Algermissen L, Delfmann P, Knackstedt R (2002) Referenzmodellierung (Reference modeling). Das Wirtschaftsstudium (WISU) 30(11):1392–1395

Becker J, Holten R, Knackstedt R, Niehaves B (2004) Epistemologische Positionierungen in der Wirtschaftsinformatik am Beispiel einer konsensorientierten Informationsmodellierung

(Epistemological positionings in IS using the example of a consensus-oriented information modeling). In: Frank U (ed) Wissenschaftstheorie in Ökonomie und Wirtschaftsinformatik (The Theory of Science in Economics and IS). Deutscher Universitäts-Verlag, Wiesbaden, pp 335–366

Beiersdorf (2009) Annual report 2009. Beiersdorf AG, Hamburg

Burnett K, Ng KB, Park S (1999) A comparison of the two traditions of metadata development. J Am Soc Inf Sci 50(13):1209–1217. doi:10.1002/(SICI)1097-4571(1999)50:13<1209:AID-ASI6>3.3.CO;2-P

Butler D (2007) MDM as a foundation for SOA. Oracle Corporation, Redwood Shores

Deloitte (2009) Telecommunications predictions—TMT trends 2009. Deloitte Touche Tohmatsu, London

Dreibelbis A, Hechler E, Milman I, Oberhofer M, van Run P, Wolfson D (2008) Enterprise master data management: an SOA approach to managing core information. Pearson Education, Boston

English LP (1999) Improving data warehouse and business information quality, 1st edn. Wiley, New York

Eppler MJ, Helfert M (2004) A classification and analysis of data quality costs. In: Chengalur-Smith S, Long J, Raschid L, Seko C (eds) Proceedings of the 9th international conference on information quality, Cambridge, pp 311–325

Fettke P, Loos P (2004) Referenzmodellierungsforschung (Reference modeling research). Wirtschaftsinformatik 46(5):331–340

Fettke P, Loos P (2007) Perspectives on reference modeling. In: Fettke P, Loos P (eds) Reference modeling for business systems analysis. Idea Group, Hershey, pp 1–21

Frank U (2007) Evaluation of reference models. In: Fettke P, Loos P (eds) Reference modeling for business systems analysis. Idea Group, Hershey, pp 118–139

Frank U, Schauer C, Wigand RT (2008) Different paths of development of two information systems communities: a comparative study based on peer interviews. Commun AIS 22(21):319–412

Friedman T (2009) Q&A: common questions on data integration and data quality from Gartner's MDM summit. Gartner, Inc, Stamford

Gregor S (2006) The nature of theory in information systems. MIS Q 30(3):611–642

Heilig L, Karch S, Pfennig R, Hofmann C (2006) SAP NetWeaver master data management. SAP Press, Bonn

Heinrich B, Henneberger M, Leist S, Zellner G (2009) The process map as an instrument to standardize processes: design and application at a financial service provider. Inf Syst e-Bus Manage J 7(1):81–102. doi:10.1007/s10257-007-0064-y

Hevner AR, March ST, Park J, Ram S (2004) Design science in information system research. MIS Q 28(1):75–105

Hüner K, Schierning A, Otto B, Österle H (2011) Product data quality in supply chains: the case of Beiersdorf. Electron Mark 21(2):141–154. doi:10.1007/s12525-011-0059-x

IBM (2005) IBM completes acquisition of Ascential software. IBM Corporation, Armonk

IBM (2006) IBM information server version 8.0: information server introduction. IBM Corporation, Armonk

Karel R (2006) Introducing master data management. Forester Research, Cambridge

Khatri V, Brown CV (2010) Designing data governance. Commun ACM 53(1):148–152. doi:10.1145/1629175.1629210

Kokemüller J (2009) Stammdatenmanagementsysteme 2009 (Master data management systems 2009). Fraunhofer-Verlag, Stuttgart

Lee YW, Pipino LL, Funk JD, Wang RY (2006) Journey to data quality. MIT Press, Cambridge

Lehmann H (2003) An object-oriented architecture model for international information systems? J Glob Inf Manage 11(3):1–18

Leser U, Naumann F (2007) Informationsintegration: Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen (Information integration: architectures and methods for integrating distributed and heterogeneous data sources). dpunkt, Heidelberg

Loshin D (2001) Enterprise knowledge management: the data quality approach. Morgan Kaufmann, San Diego

Loshin D (2008) Master data management. Morgan Kaufmann, Burlington

MA DA (2009) The DAMA guide to the data management body of knowledge. Technics Publications, Bradley Beach

March ST, Smith GF (1995) Design and natural science research on information technology. Decis Support Syst 15:251–266

Marco D (2000) Building and managing the meta data repository. A full lifecycle guide. Wiley, New York

McCann D (2010) If you build it…. CFO 26(3):29–31

Morgan DL, Krueger RA (1993) When to use focus groups and why? In: Morgan DL (ed) Successful focus groups. Sage, Newbury Park, pp 3–19

Nunamaker JF, Chen M, Purdin TDM (1991) Systems development in information systems research. J Manage Inf Syst 7(3):89–106

Otto B, Hüner K (2009) Functional reference architecture for corporate master data management. Institute of Information Management, University of St. Gallen, St. Gallen

Otto B, Hüner K, Österle H (2009) Unternehmensweite Stammdatenqualität (Corporate master data quality). ERP Manage 2009(3):19–21

Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S (2008) A design science research methodology for information systems research. J Manage Inf Syst 24(3):45–77. doi:10.2753/MIS0742-1222240302

Price R, Shanks G (2005) A semiotic information quality framework: development and comparative analysis. J Inf Technol 20(2):88–102. doi:10.1057/palgrave.jit.2000038

Pula EN, Stone M, Foss B (2003) Customer data management in practice: an insurance case study. J Database Mark 10(4):327–341. doi:10.1057/palgrave.jdm.3240122

Radcliffe J, White A (2009) Key issues for master data management, 2009. Gartner Research, Stamford

Redman TC (1996) Data quality for the information age. Artech House, Boston

Reid R, Fraser-King G, Schwaderer WD (2007) Data lifecycles: managing data for strategic advantage. Wiley, Chichester

Rosemann M, Schütte R (1999) Multiperspektivische Referenzmodellierung (Multi-perspective Reference Modeling). In: Becker J, Rosemann M, Schütte R (eds) Referenzmodellierung—State-of-the-Art und Entwicklungsperspektiven (Reference Modeling—State of the Art and Development Perspectives). Physica, Heidelberg, pp 22–44

Saaksvuori A, Immonen A (2008) Product lifecycle management, 3rd edn. Springer, Berlin

SAP (2007a) SAP NetWeaver—technology map edition 2007. SAP AG. http://www50.sap.com/businessmaps/004F8486F21E4522943507897144041B.htm. Accessed 19 Jan 2011

SAP (2007b) SAP to acquire business objects in friendly takeover. SAP AG, Walldorf

SAP (2008) SAP library—SAP NetWeaver master data management (MDM). SAP. http://help.sap.com/saphelp_mdm550/helpdata/en/43/D7AED5058201B4E10000000A11466F/frameset.htm. Accessed 19 Feb 2009

SAP (2011) SAP ERP—SAP solution map. SAP AG. http://www.sap.com/global/scripts/jump_frame.epx?content=/businessmaps/82087E7A8EB94E1F885C5A1F62C789FB.htm&CloseLabel=Fenster schliessen. Accessed 8 Feb 2011

Scheer A-W (1992) Architecture of integrated information systems—foundations of enterprise modelling. Springer, Berlin

Scheer A-W (1997) Wirtschaftsinformatik: Referenzmodelle für industrielle Geschäftsprozesse (Information systems: reference models for industrial business processes), 2nd edn. Springer, Berlin

Scheer A-W (2001) ARIS—Modellierungsmethoden, Metamodelle, Anwendungen (ARIS—Modeling Methods, Meta-models, Applications). Springer, Berlin

Scheer A-W, Thomas O, Adam O (2005) Process modeling using event-driven process chains. In: Dumas M, van der Aalst W, ten Hofstede AHM (eds) Process-aware information systems: bridging people and software through process technology. Wiley, Hoboken, pp 119–145

Schelp J, Winter R (2007) Towards a methodology for service construction. In: Proceedings of the 40th Hawaii international conference on systems sciences (HICSS-40). Los Alamitos. doi:10.1109/HICSS.2007.565

Schemm JW (2008) Zwischenbetriebliches Stammdatenmanagement: Lösungen für die Datensynchronisation zwischen Handel und Konsumgüterindustrie (Cross-company master data management: Solutions for data synchronisation between retailers and consumer goods manufacturers). Springer, Berlin

Schütte R (1998) Grundsätze ordnungsmässiger Referenzmodellierung: Konstruktion konfigurations- und anpassungsorientierter Modelle (Guidelines of reference modeling: Design of configuration- and adaptation-oriented models). Gabler, Wiesbaden

Shankaranarayanan G, Cai Y (2006) Supporting data quality management in decision-making. Decis Support Syst 42(1):302–317. doi:10.1016/j.dss.2004.12.006

Simon HA (1998) The sciences of the artificial. MIT Press, Cambridge

Sinha PR, Whitman LE, Malzahn D (2004) Methodology to mitigate supplier risk in an aerospace supply chain. Supply Chain Manage Int J 9(2):154–168. doi:10.1108/13598540410527051

Smith HA, McKeen JD (2008) Developments in practice XXX: master data management: salvation or snake oil? Commun AIS 23(4):63–72

TIBCO (2008) Managing master data with TIBCO collaborative information manager: a technical overview. TIBCO software Inc., Palo Alto

Tozer G (1999) Metadata management for information control and business success. Artech House, Boston

vom Brocke J (2007) Design principles for reference modeling: reusing information models by means of aggregation, specialisation, instantiation, and analogy. In: Fettke P, Loos P (eds) Reference modeling for business systems analysis. Idea Group, Hershey, pp 47–75

Wand Y, Wang RY (1996) Anchoring data quality dimensions in ontological foundations. Commun ACM 39(11):86–95. doi:10.1145/240455.240479

Wang RY (1998) A product perspective on total data quality management. Commun ACM 41(2):58–65. doi:10.1145/269012.269022

Wang RY, Strong DM (1996) Beyond accuracy: what data quality means to data consumers. J Manage Inf Syst 12(4):5–34

Weber K, Otto B, Österle H (2009) One size does not fit all—a contingency approach to data governance. ACM J Data Inf Q 1(1):Art. 4

White A (2008) Governance of master data starts with the master data life cycle. Gartner Research, Stamford

White A, Radcliffe J (2008) Vendor guide: master data management. Gartner Research, Stamford

White A, Newman D, Logan D, Radcliffe J (2006) Mastering master data management. Gartner Research, Stamford

Yen BK (2004) Pulling it all together. Financ Exec 20(6):60–62