A Reference Process Model for Master Data Management

Andreas Reichert, Boris Otto, and Hubert Österle

University of St. Gallen, Institute of Information Management, St. Gallen, Switzerland
{andreas.reichert, boris.otto, hubert.oesterle}@unisg.ch

Abstract. The management of master data (MDM) plays an important role for companies in responding to a number of business drivers such as regulatory compliance and efficient reporting. With the understanding of MDM’s impact on the business drivers companies are today in the process of organizing MDM on corporate level. While managing master data is an organizational task that cannot be encountered by simply implementing a software system, business processes are necessary to meet the challenges efficiently. This paper describes the design process of a reference process model for MDM. The model design process spanned several iterations comprising multiple design and evaluation cycles, including the model’s application in three participative case studies. Practitioners may use the reference model as an instrument for the analysis and design of MDM processes. From a scientific perspective, the reference model is a design artifact that represents an abstraction of processes in the field of MDM.

Keywords: Master Data Management, MDM, design science research, reference process model

1 Introduction

1.1 Motivation and Problem Statement

Master data describes the essential business entities of a company, such as supplier, customer, and product data [1-3]. The management of master data plays an important role for companies in responding to a number of business drivers such as regulatory compliance, efficient reporting in the sense of a “single version of the truth” [4-6], and the demand for having a 360 degree view on the customer [7], [8]. Therefore, master data management (MDM) is an application-independent process for the description, ownership, and management of core business data entities [3], [9].

In the past, MDM initiatives were often technology-driven neglecting the organizational aspects of the topic. With the understanding of MDM’s impact on the business drivers, many companies are today in the process of organizing MDM on a corporate-wide level [10]. In doing so, they encounter open questions. If MDM is an application-independent process, which processes need to be defined or supported by MDM? When setting up an organization, which tasks (processes) should this organization cover?
Existing literature hardly delivers any answers to these questions. Existing reference models address these questions from different perspectives. Batini and Scannapieco [11] describe activities assuring high data quality from a data management perspective (methods defining data quality metrics and improvements). Otto et al. [12] define a reference model based on six design areas defining the functional requirements of data management software. DAMA [9] structures data management activities into ten layers (architecture management, security management, quality management, etc.) and groups these layers into four activities – planning, controlling, development and operations. Governance models [2] and [13] describe responsibilities of MDM from a role perspective and their organizational integration. All these reference models focus on specific topics in the area of data management (e.g. data quality, software functionality, related roles) without addressing data management from a process perspective. Process models or descriptions, however, related to MDM are missing. A question raised by a master data manager at Beiersdorf, a Germany-based global consumer products company, illustrates the motivation: “At present, we are evaluating the tasks [of the MDM organization] and the related staffing. We created a profile of the tasks and would be obliged if you could support us with some benchmarking. Could you potentially help us out with data on […] such MDM tasks in companies of similar size and in comparable industries?” [14].

1.2 Research Question and Goal

Against this background, the article puts up the following research question: How can the tasks and activities of managing master data be structured from a process perspective? To answer this question, the paper follows the principles of Design Science Research (DSR) [15] and [16] in order to design and evaluate a reference process model for MDM. In general, a reference model is an information model that can be used in various contexts [17], [18]. 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 [19], [20]. From an epistemological perspective, the reference process model for MDM is an artifact and, thus, the result of design oriented research [21], [22]. DSR aims at designing artifacts according to scientific principles in order to be able to solve practical problems [13], [21].

2 Theoretical Background

2.1 Master Data Management

Master data specifies the essential business entities a company’s business activities are based on. Such entities are, for example, customers, suppliers, products, or employees [3]. Basically, master data can be differentiated by three concepts: master data class, master data attribute, and master data object [2]. A master data object represents an instance of a master data class (e.g. product master data) as concrete business object (a product manufactured in a certain plant at a certain point in time), and it
specifies selected characteristics of this business object (color, features, price) by means of attributes. Master data management comprises all activities for creating, modifying or deleting a master data class, a master data attribute, or a master data object [3], e.g. the modeling, provision, quality management, maintenance, and archiving of master data. All these activities aim at providing master data of good quality (e.g. master data that is complete, accurate, timely, and well-structured) for being used in business processes [2], [23].

2.2 Data Governance

A standard definition of the term “Data Governance” can be found neither in the research community nor in the practitioners’ community dealing with information systems. However, proposals to define the term have in common that data governance refers to the allocation of decision-making rights and related duties in the management of data in enterprises. According to Weber et al. [24], data governance specifies a structural framework for decision-making rights and responsibilities regarding the use of data in an enterprise. Khatri and Brown [25] see data governance referring to the assignment of decision-making rights with regard to an enterprise’s “data assets”. According to ISO [26], governance is defined as “the system by which organizations are directed and controlled”. It includes the strategy and policies within an organization which affects the management of master data.

2.3 Business Process Management (BPM)

BPM is a process oriented management method defined as „supporting business processes using methods, techniques, and software to design, enact, control and analyze operational processes involving humans, organizations, applications, documents, and other sources of information“ [27]. According to Rosemann and vom Brocke [28], each business processes is characterized by the generic BPM life cycle, namely process design and modeling, process enactment, process control and measurement, and process improvement and innovation. Process design and modeling is an approach for identifying, representing, implementing, and analyzing business processes in BPM [29]. This directly relates to the data production process [30] describing the entire life cycle of a master data object from a process perspective. According to Lynne Markus and Johnson [31] business process governance is required for the success of business processes and hence, business success. Governance refers to the direction, coordination, and control of organizations. The challenge in setting up governance for business processes is to design a cost-effective structure in which the reference process model for MDM acts as enabler for setting up this governance structure.
3 Research Approach

3.1 Research Context

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 a collaborative research project "Corporate Data Quality" at the University of St. Gallen. Since 2006, researchers, together with a number of partner companies, have been developing design artifacts in the field of data governance and master data quality management. The design process follows the principles of the Design Science Research Methodology (DSRM) [16], suggesting a sequential design process comprising iterations of design and evaluation cycles [15], [32].

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 2009, aimed at identifying the problem and motivating the research. As outlined, the research described in this paper was mainly motivated by the identification of a number of challenges in the practitioners’ community. Practitioners continuously articulated the demand for support with regard to the challenges mentioned above.

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. The proposed solution was confirmed within a focus group interview (group A). The focus groups were used for definition, design, demonstration, and evaluation purposes.

The third step comprised the design activities which followed the general principles of reference modeling [18], [20], [33]. 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 including literature review. The second and third iteration were based on the results from focus group discussions (focus groups B and C). In total, 40 persons participated in the focus groups representing user companies.

The fourth step of the design process aimed at demonstrating the applicability of the reference model. Therefore, three participative case studies for demonstration purpose were used. Furthermore, the reference process model was applied in a ‘‘real life’’ context during the participative case studies [34]. The case studies were carried out between January 2010 and November 2011.

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

- Focus group evaluation in November 2011 (focus group C)
- Multi-perspective evaluation according to the guidelines proposed by Frank [18]
The sixth step includes communication activities. Both Hevner et al. [15] and Peffers et al. [16] stipulate that DSR results must be disseminated both in the practitioners’ and the scientific community. While the former will be addressed by presentations at practitioners’ conferences, the paper at hand aims at making the research available for the scientific body of knowledge. First, it describes the reference process 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](image)

### 4 Reference Model Design

#### 4.1 Design Foundations

The design of the reference process model for MDM follows the ARIS conventions for the process architecture [35], according to which processes should be hierarchically structured. The reference model presented in this paper consists of a 3-level structure (process area, main process, and process). Main processes of the reference model for MDM are grouped in process areas. A process area consists of one or more main processes, whereas each process is assigned to only one process area. Main processes themselves consist of processes. The reference process model for MDM groups processes in main processes based on their purpose-oriented and task-oriented relationships. The visual representation of the reference process model follows the principles
of process maps, which in general aim at identifying and representing similar processes, sub-processes and functions [36] in a tabular format.

4.2 Model Overview

The value chain, one possible representation of business processes, describes a chain of activities for a company in a structured manner [37]. These activities create value, consume resources, and are linked by processes. The model differentiates value-adding and support-activities. The former represent core processes having direct impact on customer value, the latter support the core processes. Österle [38] structures processes into business-, support-, and management processes. Business processes create value for process customers. Support processes enable business processes by creating the framework for execution whereas management processes coordinate the value generation by means of target systems. Based on the analysis of 38 MDM departments by Otto and Reichert [14] and the structures of Porter [37] and Österle [38] the reference process model will be structured into three process areas: strategic, governance, and operational process area. The ARIS conventions [35] as well as the models mentioned above support the reference model from a structural perspective, proposing the three layer structure [37], [38]. The analysis of Otto and Reichert [14] supports the reference model from a content perspective. The analysis gave indications about best practices in terms of activities or processes already performed by master data organizations. This information was reflected within the design process and the focus group discussions being a basic input. In total, the reference process model is defined by 38 processes (three digit boxes 1.1.1 - 3.2.4) which are categorized into seven main processes (two digit boxes 1.1 - 3.2) being clustered in the three process areas stated above (one digit boxes 1 - 3); see figure 2. The structured was designed within the focus group discussions identifying activities performed by the participants (management level within companies; all related to MDM activities) and continuously reviewed and adapted within the DSR process. The following listing describes the three process areas.

**Strategy.** The strategic process area defines the mid- and long-term goals of MDM. Aligned with a company strategy the MDM strategy needs to define the vision and roadmap for achieving the strategic targets for MDM. The strategic process area corresponds to the management processes aforementioned.

**Governance.** The governance process area defines the standards for the operational activities related to master data management. Major activities comprise the definition of the data life cycles, the definition of quality standards and measurement metrics as well as the design of the data model and architecture. This area corresponds to the support processes aforementioned.

**Operations.** The operational process area performs the actual data life cycle assuring compliance with the standards defined within the governance process area. The data
life cycle can be seen as the core process of data management and therefore corresponds to the business process stated above.

**Main Process Level in Detail.** The three process areas are detailed by seven main processes. Each main process is detailed by processes. Exemplified, main process “Data Quality Assurance” (2.2), as part of the process area “Governance” (2) consists of 5 processes (2.2.1 – 2.2.5). Starting with the identification of a business issue (e.g. wrong e-mails to customers and high efforts in rework for correction of the e-mail-address), this might lead to preventive actions within the data creation process (e.g. additional verification activities within the sales and marketing department) as well as to the definition of measurement metrics (e.g. number of returned failure mails due to wrong e-mail-address). Furthermore, it needs to be defined what quality level the company wants to achieve (percentage of correctly delivered e-mails) and who needs to be informed about the quality level of the identified metric. The actual doing of quality measurement is part of the operational process area (3) within the process “Data Support” (process 3.2.4 “Monitor & report data quality”).

![Fig. 2. Reference Process Model for Master Data Management](image-url)
4.3 Reference Model Demonstration

Overview. Three participative case studies [34] were 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.

For demonstration of the reference process model all case studies followed the same approach applying four steps:

1. Preparation. The reference process model was not configured. For evaluation of the 38 processes, a guideline explaining and illustrating each process and a template for documentation was developed.
2. Assessment. In a workshop comprising experts related to master data management and business users, each of the 38 processes was assessed in terms of (1) is there a need for the specific process, and (2) are further processes required in order to satisfy specific master data needs. The assessment was conducted independent of existing processes.
3. As-is analysis. For each of the 38 processes the current status of realization was analyzed with regard to improvement potential or already established processes.
4. To-be concept. The analysis of activity 2 and 3 lead to the identification of processes that are demanded but are provided not at all or not in the quality desired. For each of these processes it has to be stated whether it can be implemented by means of new processes or adaptation of existing processes.

Case A. Case A is a joint venture between a private equity firm and a German industrial conglomerate founded in 2008. The company, with presence in more than 100 countries, provides communication products on an employee base of 11,000 and revenues of 3.2 bn € (~ 4.3 bn US$).

Due to several business drivers, e.g. changing business models in the past years (from hardware business to full-service orientation) and cost pressure the company initiated a global MDM project with the focus on product, customer, and vendor master data. Objectives of the project are the design of global governance, the improvement of data quality as well as global master data life cycles. In this context the company applied the reference process model following the steps described above.

Case B. Case B is an engineering company founded in 1889. The company, with presence in more than 190 countries, provides medical equipment on an employee base of 11,000 and revenues of 2.2 bn € (~2.9 bn US$) in 2011.

Due to an internal reorganization, the two major business segments (medical and safety equipment) are moving on one Enterprise-Resource-Planning-System (ERP) from a technical point of view. Organization wise, MDM for product data is managed within the Research & Development department within the medical division, whereas product master data for the safety division is managed by the Operations department. Strategic direction is given that both departments unify their activities and service portfolio. As part of these activities, both departments applied the reference process model for structuring their service portfolio.
Case C. Case C is a retail company founded in the 1930s. The company, with presence in 26 countries, operates more than 10,000 stores on an employee base of 170,000 and revenues of 42 bio € (~56.5 bio US$) in 2010.

As part of an ERP-implementation replacing the existing solution with standard software, the direction was given to set up a strategic MDM initiative. Structuring this initiative on the base of the Corporate Data Quality-framework [14] - strategic, organizational, and system perspective - the company defined its MDM activities for product- and supplier master data based on the reference process model.

Demonstration Results. Table 1 shows the related design decisions of all cases.

The application of the reference process model in the three case studies indicates that the design of the three process areas (Strategy, Governance, Operations) is appropriate in order to structure MDM functions. Furthermore, all companies agreed that all 38 processes are relevant, excluding the process “Develop and adapt vision”, which was removed in company C. The differentiation between MDM-activities and IT-activities remains complex. Whereas companies A and B removed the IT-related processes from the MDM process model and allocated them within the IT-processes, company C included these processes accepting double definitions in the company process model (e.g. implementation & testing activities within MDM- and IT-process models). A further adjustment done by all companies was the renaming of the processes within “Data Life Cycle” as all companies already have an established process model and nomenclature for these activities.
Table 1. Design Decisions for Reference Process Model for MDM

<table>
<thead>
<tr>
<th>Design Decision</th>
<th>Justification</th>
<th>Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process “Define strategic targets” removed (1.1.3)</td>
<td>• Activities integrated in process “Align with business/IT strategy”</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>• No explicit MDM strategic targets required as they should be integrated in existing target systems</td>
<td></td>
</tr>
<tr>
<td>Process “Model Workflows/UIs (User Interfaces) moved from main process “Architecture” to “Standards &amp; Guidelines” (2.4.3)</td>
<td>• Focus for activity is set on conceptual design rather than technical implementation aspects</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>• Technical implementation needs to be covered by IT-processes. Case A only covers the conceptual part of the workflow design. The implementation process will be described outside of this process</td>
<td></td>
</tr>
<tr>
<td>Process “Monitor &amp; report” (in context of Quality Assurance) moved from main process “Support” to “Quality Assurance” (3.2.4)</td>
<td>• Mix of governance and operational activities in main process “Governance”</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>• However, focus is set on end-to-end process including both aspects</td>
<td></td>
</tr>
<tr>
<td>Process “Test &amp; Implement” (in context Architecture) removed (2.4.5)</td>
<td>• Testing activities defined within IT-processes and do not need to be covered by data management processes</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>• Removal will eliminate double definitions within company</td>
<td>X</td>
</tr>
<tr>
<td>Processes of main process “Life Cycle” renamed (3.1)</td>
<td>• Naming of processes aligned with company specific naming conventions as processes were already defined</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Process “Mass data changes” added to “Support” (new 3.2.5)</td>
<td>• New process added as activity is performed on continuous base and should be covered by data management processes</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Process “Develop and adapt vision” removed (1.1.1)</td>
<td>• Company strategies not defined by visions but by strategic targets</td>
<td>X</td>
</tr>
<tr>
<td>Processes “Adapt data life cycle”, “Adapt standards and guidelines”, “User trainings”, and “Support Processes” merged to “Standards for operational processes” (2.1.2 - 2.1.6)</td>
<td>• Activities of all processes remain existing</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>• Goal is simplification of process model</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Description of all activities, which have been merged to the new process, will be created on the work description level, which will underlay the process model for execution of processes (including process flows, responsibilities, etc)</td>
<td></td>
</tr>
<tr>
<td>Processes “Test and implement (data model)” and “Roll out data model changes” removed (2.3.4 - 2.3.5)</td>
<td>• Activities defined within IT service portfolio outside of this process model</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>• As activities are already defined, they do not need to be covered within this structure</td>
<td></td>
</tr>
<tr>
<td>Main process “Data Architecture” removed (2.4)</td>
<td>• Activities defined within IT service portfolio</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>• Clear separation between business requirements and modeling of data and IT realization (integration architecture etc.)</td>
<td></td>
</tr>
<tr>
<td>Process “Data analysis” in main process “Support” added (new 3.2.6)</td>
<td>• Requests for one-time analysis of master data as service offering defined which are not covered by standard reports</td>
<td>X</td>
</tr>
</tbody>
</table>

826
5 Evaluation

With a particular focus on the evaluation of reference models, Frank [18] has proposed a framework comprising four perspectives of evaluation. This framework is used for evaluation of the reference process model presented in this paper.

**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 (one day for preparation and one day for application in the case presented). Tools supporting the processes of creating the process model can be created at low effort (Microsoft Excel based templates for documentation). Using the model does not lead to direct cost savings. However, the as-is analysis might identify processes not implemented yet or potentials for consolidation. Both the focus group interviews and the final validation by application within the case studies have shown that the reference process model is capable of simplifying exchange of knowledge. Furthermore, all companies stated that the transparent design of strategic, governance, and operational processes facilitated the management of master data related activities. The companies were able to use the model as a starting point in order to optimize the data life cycles and identify governance responsibilities required.

**Deployment Perspective.** The focus group interviews and the application of the reference process model within the three companies 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.** The model’s simple structure ensures its easy adaptability [18]. The focus group interviews indicated that a differentiation between mandatory and optional processes may be made. “Communication strategy” as important activity should be added, which has already been done within the development of the process model.

**Epistemological Perspective.** The validation by application of the reference model at the three companies has shown that the model is capable of abstracting and representing reality. Critical distance is ensured by explication of use cases. Moreover, explication of the model design process ensures that scientific principles are followed (such as verifiability and reproducibility of the artifact).

6 Conclusion, Contribution, and Outlook

The paper describes a reference process model for MDM. The design process spanned the six steps as proposed by DSRM and includes several design and evaluation cycles. The reference process model is beneficial with regard to both the advancement of the scientific state of the art and the state of the art in practice.
The reference process model supports companies trying to overcome the challenges listed in the previous sections. It helps to create a common terminology both for internal and external communication. Furthermore, it offers an instrument for evaluating existing and identifying required MDM processes. The model can identify processes not defined yet and it can be used as a starting point for the allocation of resources and responsibilities. This leads back to the definition of BPM governance [31] identifying direction, coordination, and control. The model could lay the foundation and define the scope of what needs to be coordinated and controlled.

The reference process model can make use of existing approaches by applying existing methods as part of the detailed process descriptions [11], [12]. The reference model can be seen as overarching approach for bringing MDM into organizations using a process view and linking the processes to responsibilities (data governance).

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. Explication of the research process allows verification, correction, and differentiation of this representation. Furthermore, the reference process model represents an abstraction of processes in the field of MDM. Hence, it forms a “theory for designing” [39] and contributes to the integration of governance and management of processes in the context of MDM.

The reference process model has its limitations due to its focus on the business layer of the process view of MDM [40] and due to the fact that other ARIS views and levels were not modeled. Hence, the application of the reference process model is restricted to use cases similar to the two described. Besides scientific validation of the reference process 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 organizational view offers potential for designing relevant artifacts. Based on case studies, generic characteristics of MDM organizations (roles and responsibilities, for example) could then be identified to constitute the basis for conceptualizing rights and roles as the reference process model’s organizational view. Besides, interdependencies between individual processes of the reference process model could be identified.

Acknowledgements

The research presented in this paper is a result of the Competence Center Corporate Data Quality (CC CDQ) at the Institute of Information Management at the University of St. Gallen (IWI-HSG). The CC CDQ is a consortium research project and part of the research program Business Engineering (BE HSG).

References