Institute of Information Management

University of St.Gallen

Functional Reference Architecture for Corporate Master Data Management

Boris Otto, Kai M. Hüner

Report No.: BE HSG / CC CDQ / 21
Chair: Prof. Dr. H. Österle
Version: 1.0
Date: May 31, 2009

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

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

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

Content

1 Introduction 8
1.1 Motivation and problem identification 8
1.2 Research context, research objective, and research process 10
2 Basics 13
2.1 Data, information, and knowledge 13
2.2 Master data and master data quality 14
2.3 Corporate MDM: initial situation and design areas 15
3 Functional Reference Architecture 20
3.1 Introduction 20
3.2 Master Data Lifecycle Management 20
3.2.1 Overview 20
3.2.2 Data Creation 21
3.2.3 Data Maintenance 22
3.2.4 Data Deactivation 23
3.2.5 Data Archiving 23
3.3 Metadata Management and Master Data Modeling 24
3.3.1 Overview 24
3.3.2 Data Modeling 24
3.3.3 Model Analysis 26
3.3.4 Metadata Management 26
3.4 Master Data Quality Management 28
3.4.1 Overview 28
3.4.2 Data Analysis 29
3.4.3 Data Enrichment 31
3.4.4 Data Cleansing 32
3.5 Master Data Integration 33
3.5.1 Overview 33
3.5.2 Data Import 34
3.5.3 Data Transformation 34
3.5.4 Data Export 35
3.6 Cross Functions 36

3.6.1 Overview 36
3.6.2 Automation 36
3.6.3 Reports 37
3.6.4 Search 38
3.6.5 Workflow Management 39
3.7 Administration 40
3.7.1 Overview 40
3.7.2 Data History Management 41
3.7.3 User Management 41
4 Product Analysis 43
4.1 Introduction 43
4.2 IBM InfoSphere 43
4.3 Oracle Master Data Management Suite 45
4.4 SAP NetWeaver Master Data Management 45
4.5 TIBCO Collaborative Information Manager 47
4.6 Coverage of Functional Reference Architecture 48
5 Selected Case Studies 54
5.1 SAP NetWeaver MDM at Oerlikon Textile 54
5.2 IBM Master Data Management at PostFinance 56
6 Summary and Outlook 58
Appendix A Focus Group Participants 61
A. 1 Focus group interview on November 25, 2008 61
A. 2 Focus group interview on February 9, 2009 62
A. 3 Focus group interview on February 18, 2009 63
Appendix B Functional Reference Architecture − Overview 64

List of Figures

Figure 1-1: Application scenarios of the Functional Reference Architecture 9
Figure 2-1: Terms used for describing data and data structures 14
Figure 2-2: Design areas for corporate MDM 17
Figure 3-1: Functional categories and areas 20
Figure 3-2: Functions of Master Data Lifecycle Management 21
Figure 3-3: Functions of Metadata Management and Master Data Modeling 24
Figure 3-4: Functions of Master Data Quality Management 29
Figure 3-5: Functions of Master Data Integration 33
Figure 3-6: Cross Functions 36
Figure 3-7: Functions of Administration 40
Figure A-1: Functional Reference Architecture − Overview 64

List of Abbreviations

API	Application Programming Interface
BE	Business Engineering
CC	Competence Center
CDQ	Corporate Data Quality
CDQM	Corporate Data Quality Management
CIM	Collaborative Information Manager
DQM	Data Quality Management
DRM	Data Relationship Management
DSAP	Deutschsprachige SAP-Anwender-Gruppe (German for German-speaking SAP user group)
ERP	Enterprise Resource Planning
ETIM	Elektro-technisches Informationsmodell (German electronic products trade standard)
GUI	Graphical User Interface
HSG	Hochschule St. Gallen (German for University of St. Gallen
IT	Information Technology
IWI	Institute of Information Management
MDM	Master Data Management
SAP NetWeaver MDM	SAP NetWeaver Master Data Management
SAP NetWeaver PI	SAP NetWeaver Process Integration
SOA	Service-Oriented Architecture
TIBCO CIM	TIBCO Collaborative Information Manager
XML	Extensible Markup Language
XSLT	Extensible Stylesheet Language Transformation
API	Application Programming Interface

Preliminary Remarks and Acknowledgements

The Functional Reference Architecture presented in this paper aims at supporting project managers, master data stewards, and the like in establishing quality oriented Master Data Management (MDM) in their companies. It also offers help in the process of evaluating software products, and MDM roadmap planning.

Among other things, the paper examines selected MDM products with regard to their capability to meet the functions specified in the Functional Reference Architecture. It is important to note that neither the selection of products is by any means exhaustive, nor does the paper in any way rank the products selected.

The Functional Reference Architecture for corporate MDM presented and discussed in this paper is an outcome of applied research the Institute of Information Management (IWI-HSG) of the University of St. Gallen (HSG) has been doing in the course of a research program named Business Engineering (BE HSG) since 1989. More precisely, the Functional Reference Architecture is a result developed by the Competence Center Corporate Data Quality (CC CDQ), which is one of several consortium research projects accommodated in the IWI-HSG. The research was supported by sponsorships from IBM Deutschland GmbH, Stuttgart-Vaihingen, and SAP Deutschland AG & Co. KG, Walldorf/Bd.

The authors together with the entire IWI-HSG team in the CC CDQ would like to thank the sponsors, the consortium members in the CC CDQ, the participants of the DSAG Workgroup MDM, and everybody else contributing to the work presented in this paper by providing their recommendations and ideas. Our special thank goes to Mr. Thomas Kägi for his substantial support of our work in the course of his bachelor studies at the HSG.

Summary

Master Data Management (MDM) brings about two major challenges for companies: 1) Companies need to cope with the complexity of the subject, and 2) companies see themselves confronted with a wide range of IT products and solutions for MDM. Presenting a Functional Reference Architecture for Corporate Master Data Management, the present paper identifies and describes from a business perspective functional requirements MDM software should meet. The Functional Reference Architecture provides a basic terminology, a check list, and an assessment scheme for various application scenarios, like product evaluation, roadmap planning or exchange of information and experiences. Furthermore, MDM solutions of four software providers are examined with regard to their capability to meet the functions specified in the Functional Reference Architecture.

1 Introduction

1.1 Motivation and problem identification

Today's companies not only need to cope with extremely short innovation cycles and time-to-market but also with increased complexity caused by the need to harmonize business processes on a global level. Also, time intervals for making strategic decisions are getting shorter and shorter, with data volumes such decisions need to be based on getting bigger and bigger [Kagermann/Österle 2006].

Many companies' IT departments are supposed to follow a ‚Do more with less' philosophy, aiming at working efficiently while at the same time reducing costs. As operating costs usually represent the biggest cost driver, companies tend to eliminate and/or consolidate application and infrastructure systems. When removing superfluous applications and systems or when looking for alternative, more cost-efficient software solutions, however, companies must ensure that all IT functions required to do business well are still available.

By presenting a Functional Reference Architecture the present paper aims at describing and categorizing functions deemed necessary for doing corporate MDM. From an IT perspective these functions describe requirements information systems supporting MDM should meet. The functions are deliberately described from a business perspective in order to achieve a certain level of abstraction allowing to compare different products and solutions. It is important to note that the Functional Reference Architecture does not claim to be the ultimate catalog of MDM functions but rather offers companies orientation in the process of selecting an appropriate MDM solution. At any rate, companies interested in implementing MDM should further specify the functions proposed and complement them with company specific aspects.

The reason why we propose a Functional Reference Architecture is that companies have long been asking for tools that offer orientation when dealing with MDM (see

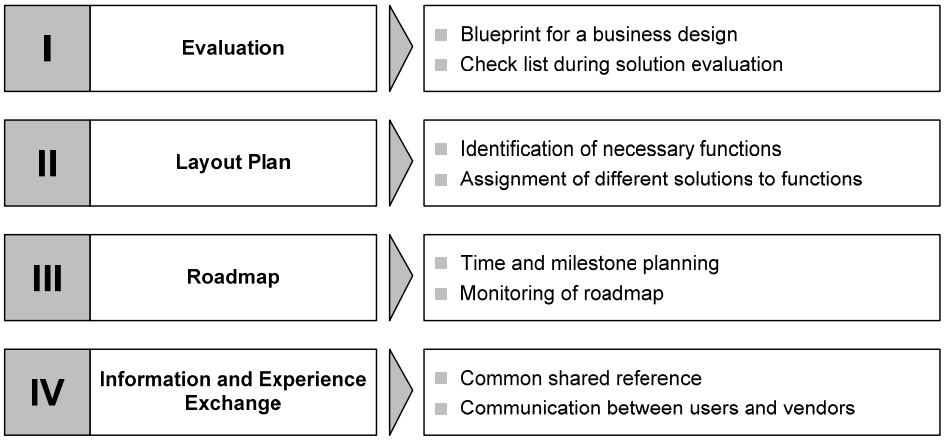


Figure 1-1: Application scenarios of the Functional Reference Architecture

Section 1.2). While the subject is characterized by considerable complexity (see Section 2.3), companies see themselves confronted also with a wide range of products and solutions offering support in meeting MDM requirements. Taking into account both of these challenges, Figure 1-1 shows four application scenarios of the functional reference architecture.

- *Evaluation*. The Functional Reference Architecture serves as a tool for evaluating software products. It is used here as a catalog from which people responsible for MDM in a company can select the functions that are needed. This selection is then presented to software providers in a form similar to specifications.
- *Layout Plan*. The Functional Reference Architecture serves as a tool for identifying activities to be supported in a companywide MDM project. It is used here as a reference model helping identify a to-be architecture (see *Roadmap* scenario). All functions to be supported must be checked as to which of them are provided by existing information systems (as-is architecture identification) and for which new software solutions are needed (see *Evaluation* scenario).
- *Roadmap* Based on the identification of the as-is architecture (see *Layout Plan* scenario), the roadmap to the to-be architecture must be defined. Only in very

rare cases is it possible to realize the complete to-be architecture in a single step. What is usually needed is a sequential introduction accompanied by process planning that is permanently monitored. The Functional Reference Architecture serves as a tool for planning and monitoring the process (e.g. by allowing people responsible for MDM to visualize the process from as-is to to-be architecture by marking functions already supported).

- *Information and Experience Exchange*. Following the principles of a reference model, the Functional Reference Architecture serves as a tool specifying basic terminology. While manufacturers and providers of software are required to describe their products in a way that allows comparison of products (which is beneficial for customers), customer requirements can be specified more clearly and met more efficiently (which is beneficial for manufacturers and providers).

1.2 Research context, research objective, and research process

The Functional Reference Architecture for corporate MDM presented and discussed in this paper is an outcome of applied research the Institute of Information Management (IWI-HSG[1]), University of St Gallen (HSG[2]) has been doing in the course of a research program BE HSG[3] since 1989. More precisely, the Functional Reference Architecture is a result developed by the CC CDQ[4], which is a consortium research project [Back et al. 2007, Otto/Österle 2009], in the course of which artifacts (e.g. architectures, methods, reference models) aiming at solving problems in Corporate Data Quality Management (CDQM) are being constructed and evaluated.

Following the principles of Design Science Research, the research objective behind this paper is (as already outlined in Section 1.1) the construction of a Functional Ref-

[1] Website IWI-HSG: http://www.iwi.unisg.ch
[2] Website HSG: http://www.unisg.ch
[3] Website BE HSG: http://www.iwi.unisg.ch/behsg/
[4] Website CC CDQ: http://cdq.iwi.unisg.ch

erence Architecture, which can be used by companies in scenarios like the ones described. The research process can be divided into four phases:

- *Phase I – Basis for discussion*. Based on the experiences made in the CC CDQ a first proposal for a functional scheme was put up, taking into consideration similar approaches [Heilig et al. 2007, DAMA 2008, Dreibelbis et al. 2008, Mosley 2008, White/Radcliffe 2008] as well as findings from bilateral projects, workshops, and conferences. The result of this phase was a rough structural scheme subdividing MDM into six areas (lifecycle management, quality management etc.) and a list of tasks for each area (create master data, cleanse master data etc.).
- *Phase II – Discussion and further specification*. The structural scheme was discussed by 34 experts in three focus groups (Appendix A contains a list of the participants). Adaptations proposed were iteratively integrated into the concept. The result of this phase was a Functional Reference Architecture comprising three structural levels: functional categories, functional areas, and functions (see Section 3).
- *Phase III – Product analysis*. MDM products of four software manufacturers were examined against the Functional Reference Architecture, and specific functions and components of the products have been classified by the architecture (see Section 4). The result of this phase is the product analysis overview presented in Section 4.6.
- *Phase IV – Review*. Following the documentation of the Functional Reference Architecture and the product analysis conducted, the software manufacturers and the focus group participants had the opportunity to review the results and propose further adaptations. Sections 3 and 4 present the Functional Reference Architecture and the product analysis after integration of all final propositions from both experts and manufacturers.

Prior to the construction of the Functional Reference Architecture we asked users of various MDM products to evaluate existing functionality and help identify missing functionality of these solutions. However, it was not possible to gain valid and reliable data from this survey, firstly because there was no common understanding of MDM tasks and functions among interviewees, and secondly because the MDM products were so different that it was hardly possible to compare them with one another. That experience additionally contributed to pursuing the goal to construct a Functional Reference Architecture that is supposed to offer – among other things – a basic terminology to be used for market and product analysis.

2 Basics

2.1 Data, information, and knowledge

Data are made available by information systems in a certain context (for example, customer data used in a business process). When data are used by a human being, they turn into information, and information finally turns into knowledge. A detailed discussion on the differentiation of data, information, and knowledge is given by BOISOT & CANALS [2004], SPIEGLER [2000, 2003], DAVENPORT & PRUSAK [1998] und BOURDREAU & COUILLARD [1999]. For the present paper we postulate some simplifying key assumptions. Data store and describe attributes of objects and processes from the real world. By processing data (e.g. by analyzing, interpreting, or structuring them), data turn into information. This transformation is usually done by computer programs (by interpreting a certain sequence of characters as a data of birth, for example). So, while any transformation of data into information usually is independent of the user, it is by any means dependent on the context the data are used (the computer program, in our example) [Tuomi 1999]. Knowledge, finally, is generated by processing information (by linking, qualifying, quantifying, or disseminating it, for example). *This* transformation does depend on the user and their specific situation. The knowledge resulting from this transformational process allows users to respond to certain events. For example, in a machine maintenance process a figure (piece of data) is interpreted as a certain value of a metric (piece of information) triggering a maintenance activity (action), because a certain threshold value (existing knowledge) has been exceeded (event).

Regardless of such clear theoretical differentiation between data and information, practitioners use the term ‚data' in a broader sense. Master data (e.g. customer or material master data) are not just values (e.g. 0721) but comprise also the act of interpreting by means of a certain scheme (here: a telephone area code) or in a certain context (here: a customer's phone number). As the functional reference architecture

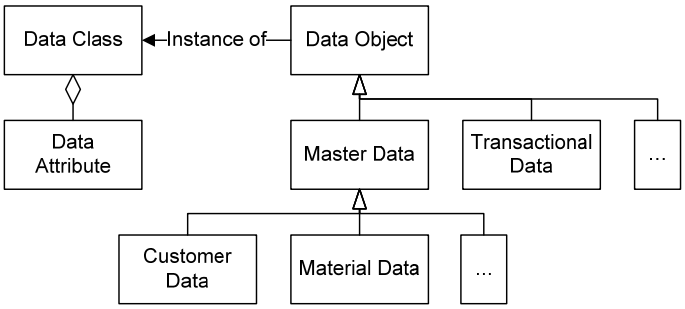


Figure 2-1: Terms used for describing data and data structures

to be presented in this paper does not so much aim at a theoretical differentiation of certain terms but rather focuses on the practical use of information systems for MDM, we favor a broader use of the term 'data', as it is common in other research contexts too [Pipino et al. 2002]. Figure 2-1 shows the terms used in this paper for describing data, data structures, and the relations between them.

- *Data Class*. Data classes are structures consisting of one or more data attributes. For example, the way customer data are structured (i.e. attributes and relations) defines how a company does customer data management.
- *Data Attribute*. Data attributes describe concrete aspects of data classes (a customer's date of birth, for example). Data attributes are defined by a denominator and a data type.
- *Data Object*. Data objects are instances of data classes (data about a certain customer, for example). Data classes are instantiated by assigning values (a sequence of numbers, for example) to data attributes (the telephone number attribute of the customer data class, for example). Data objects of one data class constitute a data set (customer or material data, for example).

2.2 Master data and master data quality

Master data describe basic entities which a company's business activities are based on. Entities are, for example, a company's business partners (customers or suppliers), products, or staff [Mertens et al. 2004]. Basically, master data differ from other types of data in four ways:

- Unlike transaction data (e.g. invoices, orders, or delivery notes) and inventory data (e.g. stock on hand, account data), master data describe always basic characteristics (e.g. the age, height, or weight) of objects from the real world.
- Pieces of master data usually remain largely unaltered. For example, as the characteristic features of a certain material are always the same, there is no need to change the respective master data. And while during the lifecycle of a product various attribute values are added over time (dimensions and weight, replenishment times etc.), the basic data remain unaffected.
- Instances of master data classes (data on a certain customer, for example) are quite constant with regard to volume, at least when compared to transaction data (e.g. invoices or purchase orders).
- Master data constitute a reference for transaction data. While a purchase order always involves the respective material and supplier master data, the latter do not need any transaction data in order to exist.

As ‚good' master data always meet a certain purpose defined by the user, data quality often is seen under a 'fitness for use' aspect. A more concrete determination of data quality is based on various data quality dimensions, like consistency or completeness [Redman 1996, Wang/Strong 1996, English 1999] (with consistency indicating the degree to which data values are consistent across a number of data sources, and completeness indicating the degree to which the total of data is identified and captured).

2.3 Corporate MDM: initial situation and design areas

High-quality master data are a central prerequisite for companies in order to be able to perform as desired. Companies cannot react properly to business drivers such as legal provisions, integrated customer management, effective reporting, or business process harmonization, if their master data are inconsistent, incomplete, incorrect,

not up-to-date, or not available. Despite its importance for doing business properly and professionally, MDM is still being largely neglected in many companies.

- The whole matter of dealing with master data is often not seen as a discrete field that requires central management but rather as a subordinated task that can be executed by different roles in the company. Typically, persons responsible for certain business processes deal with the process output and the activities relevant for the process. Master data are taken into consideration only as long as they are in the scope of the particular process. The fact that the same master data are being used – and perhaps modified – in other business processes usually is not given any importance. Similarly, persons responsible for certain application systems see master data only within the boundaries of these applications. The fact that the same master data are used by other application systems again is not taken into account.
- MDM is a complex issue. The majority of a company's master data is used throughout the whole organization. Single classes of master data have relations with single units, divisions, departments, regional branches, functions, or business processes of the company. Only persons with long-standing experience in the company (and in the industry the company is in) can be expected to be capable of understanding these complex relationships sufficiently. Yet usually there are not many persons in an organization who have this experience *and* the verbal capabilities needed to communicate the issue of MDM to all relevant stakeholders in the company (management, IT department, functional departments etc.).
- MDM cannot be done alone either by a company's IT department or by individual functional departments. While the meaning of the specific entities (such as customers, suppliers, or materials) in combination with pertinent attributes (such as addresses, trade register numbers, or product group codes)

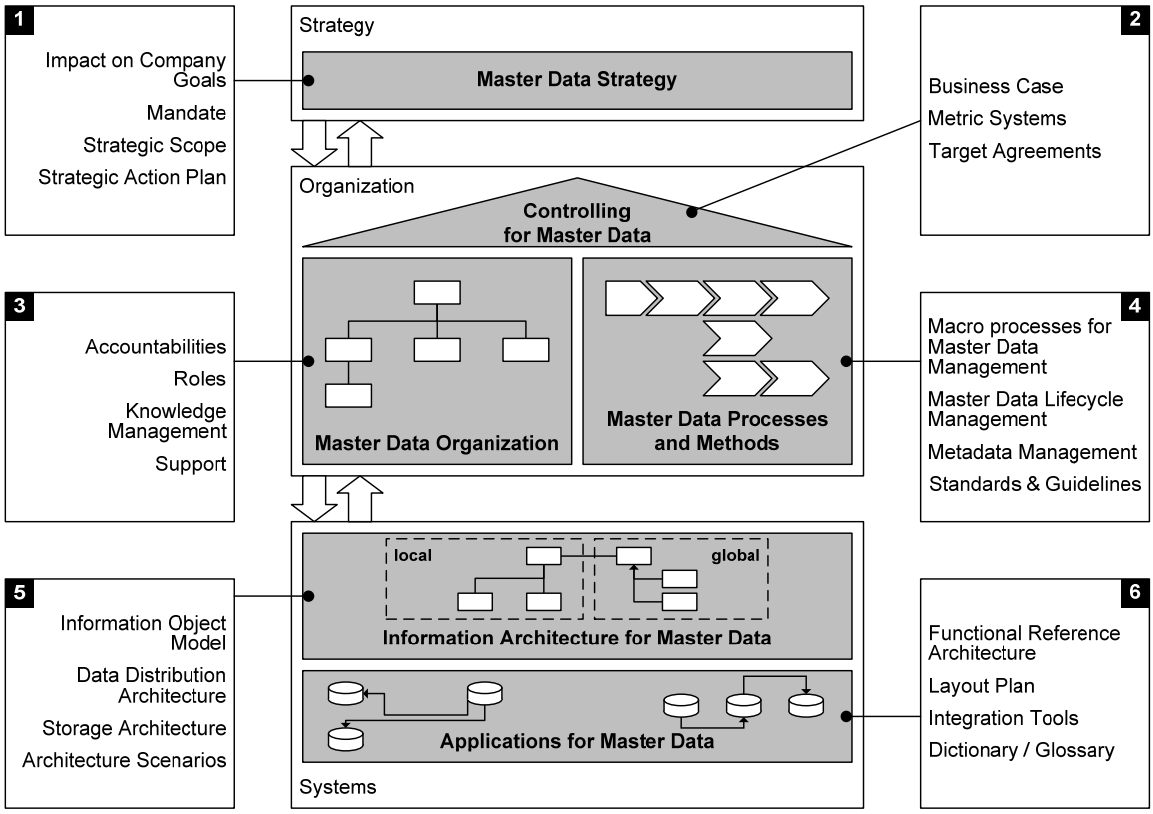


Figure 2-2: Design areas for corporate MDM

can only be assessed properly by people from the functional departments, IT experts are needed to plan, construct, and operate information systems representing these entities in master data objects.

Figure 2-2 shows design areas for corporate MDM following the principles of Business Engineering [Österle/Winter 2003]. Business Engineering is a scientific method developed by the Institute of Information Management of the University of St. Gallen, allowing to design business transformations that are based on the strategic use of IT. The guiding principle of Business Engineering is that such business transformations are to be designed on three different levels, namely the strategic level, the organizational level, and the system level. The design areas for corporate MDM are specified as follows:

- *Master Data Strategy*. As MDM is affected by various business drivers (risk management, compliance, business process integration and standardization

etc.), it must be considered a company-wide endeavor. Requirements resulting from legal provisions, integrated customer management, or reporting have an effect on the company as a whole, not just on single units, divisions, departments, or regional branches. Thus MDM per se is of strategic relevance.

- *Controlling for Master Data*. Controlling for Master Data is responsible for identifying the as-is situation prior to the establishment of corporate MDM and for translating the Master Data Strategy into concrete objectives. Components of Controlling for Master Data are a business case specifying MDM measures planned, metric systems for assessing master data quality, and objectives for target agreements.
- *Master Data Organization*. As MDM is vital for a company as a whole, it must be coordinated across a company's units, divisions, departments, or regional branches. Many companies have a virtual Master Data Organization with workers remaining in their original reporting lines while additionally being integrated in a certain MDM specific reporting line.
- *Master Data Processes and Methods*. In order to handle master data properly and in a standardized way across the entire organization and to ensure master data quality, standard procedures and guidelines must be embedded in company's daily processes. Such standard procedures and guidelines should refer both to business processes, in which workers perform activities related to their line, and project work.
- *Master Data Information Architecture*. While master data are important for business processes and have far-reaching organizational implications, in the end it all comes down to data stored in and exchanged between information systems. In many organizations the design and maintenance of these relationships is no trivial thing. As organizations often are quite complex and information system landscapes have grown enormously over time, there is often

little or no transparency at all with regard to master data storage, distribution, and interpretation.

- *Master Data Application Systems*. What application systems are to be used for MDM must clearly be specified. The functional reference architecture presented in this paper is supposed to offer support in this specific design area (see Section 1.1).

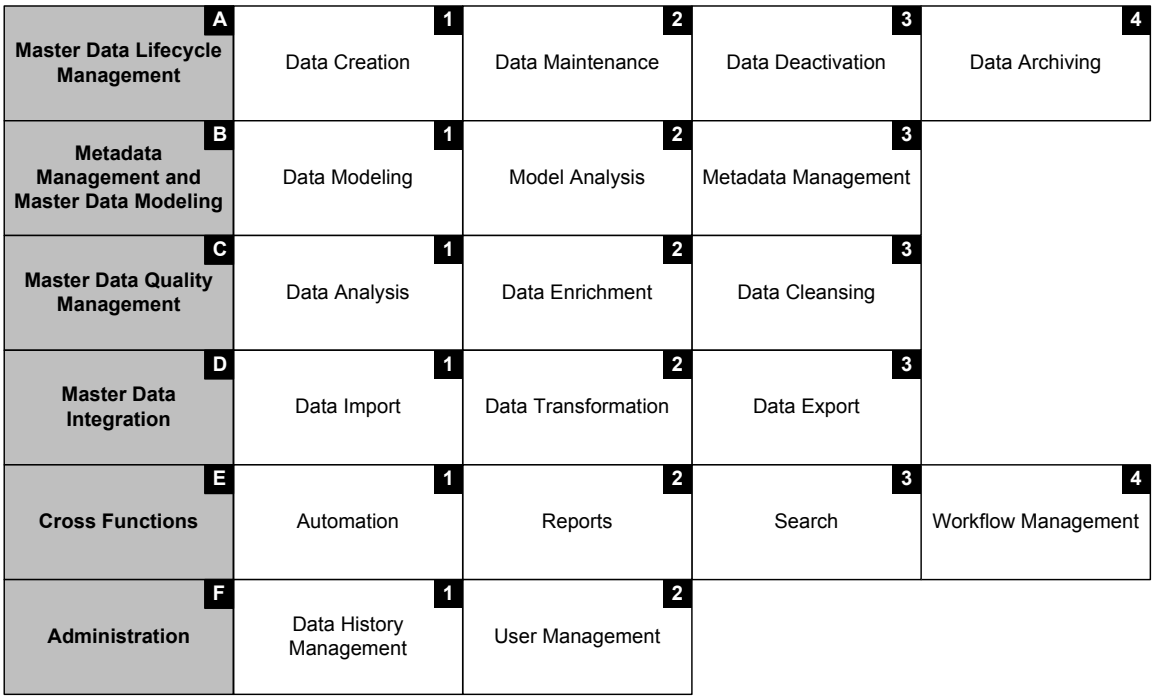


Figure 3-1: Functional categories and areas

3 Functional Reference Architecture

3.1 Introduction

The following sections outline and explain the elements of the functional reference architecture presented in this paper. The functional reference architecture is subdivided into three levels. Figure 3-1 shows the first level, specifying six functional categories, and the second level, specifying 19 functional areas. The third level, which is not shown in the image, finally specifies 72 discrete functions, which constitute a functional reference architecture that can be used to analyze MDM products (see Section 4). Functions that can be used in several areas (e.g. *Bulk Editing*) are explained only once.

3.2 Master Data Lifecycle Management

3.2.1 Overview

A master data object's lifecycle starts with its creation during business operations and ends with its deactivation and/or archiving [Redman 1996]. *Master Data Lifecycle*

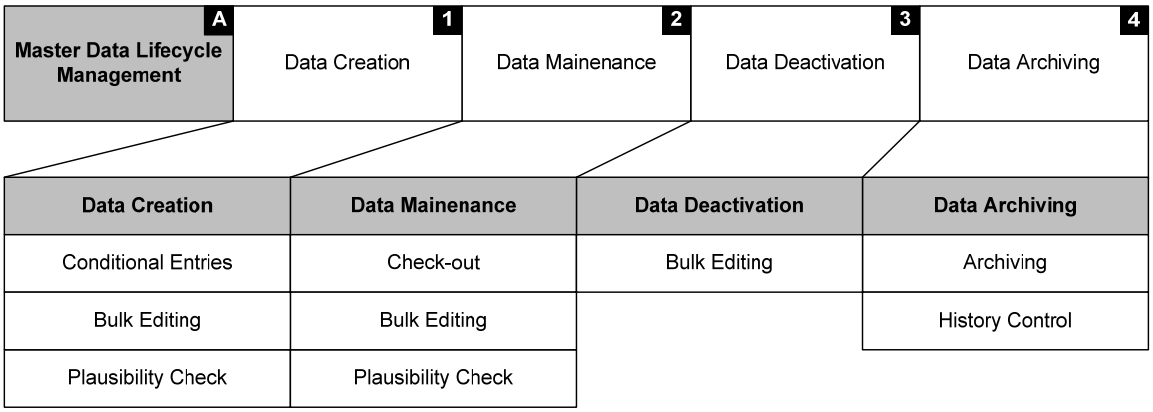


Figure 3-2: Functions of Master Data Lifecycle Management

Management describes all activities data users or data managers do with master data during their entire lifespan [Lee et al. 2006]. Figure 3-2 shows the functional areas and the functions of this category. Self-explaining functions, such as *Creation* or *Deactivation*, are not stated explicitly.

> *Any measures supposed to ensure data quality should be placed along the entire data lifecycle. Time alone can have a negative effect on data quality, for example when address data gets old.*
>
> *Ralf Jäger, Client Vela GmbH*

3.2.2 Data Creation

Conditional Entries

By means of *Conditional Entries* relations between master data classes that change depending on values of the associated classes, can be efficiently modeled and stored. An example for such a relation would be the relation between customer master data and supplier master data, e.g. terms and conditions with suppliers, specifying different discount rates for different purchase quantities.

Bulk Editing

See *Bulk Editing* (*3.2.3 Data Maintenance*). As opposed to data maintenance, in the process of data creation bulk editing usually affects only certain attributes (e.g. the zip code of a certain customer group).

Plausibility Check

Plausibility Check ensures that no invalid data are entered in application systems. The function could use reference lists that contain correct addresses, correct names etc.

> *Data clearance and data consolidation must take place at the beginning of a workflow. It must not be a subsequent, separate measure to be done ex post by a central 'master data authority'. Data clearance and data consolidation must be integrated into the data lifecycle at the beginning of the process chain, and it must be ensured by using intelligent search programs and by designing workflows accordingly from the outset.*
>
> *Karlheinz Sturm, Voith Turbo GmbH & Co.KG, Heidenheim*

> *During the process of data creation an MDM tool should ensure that no invalid data are entered in application systems. Such a tool could use reference lists allowing to verify, for example, whether a certain name stands for a male or a female person or whether a certain address really exists. Plausibility rules configured by users could prevent other errors frequently occurring during the data creation process, such as confusing gross weight with net weight. A good tool should come standard with such reference lists and plausibility rules.*
>
> *Detlef J. Königs, Mars Services GmbH*

3.2.3 Data Maintenance

Check-out

Check-out prevents data objects from being edited by other users. Usually, data which are temporarily checked out can be read by other users, but these users cannot alter attribute values during this time. When the editing of data objects in the check-out mode is complete, the data are checked in again.

Bulk Editing

Bulk Editing allows to edit a number of data objects at a single time (e.g. by ticking check boxes in a list), i.e. an editing process does not have to be executed for single data objects individually.

Plausibility Check

See *Plausibility Check* (*3.2.2 Data Creation*).

3.2.4 Data Deactivation

Bulk Editing

See *Bulk Editing* (*3.2.3 Data Maintenance*).

> *Deactivation of master data is one of the most complex issues in master data management, as each class of master data has different requirements regarding differing use context and possibly shared use. The basis of each process of deactivation is always the use of a data object. One can distinguish the scenarios a) deactivation because a data object no longer exists, b) immediately deactivation due to legal, financial, or other extraordinary reasons, and c) deactivation of a duplicate. Pre-levels of final deactivation are lock or deletion flags to block certain user activities.*
>
> *Helge Enenkel, Voith Paper Holding GmbH & Co. KG*

3.2.5 Data Archiving

Archiving

Archiving allows to persistently store data objects for a defined period of time. This function supports compliance with relevant legal provisions[1].

History Control

History Control allows to archive different versions of any piece of master data. As data must not be overwritten when being archived, this function ensures that any data object can be reconstructed as it was at a certain point in time.

> *History control should be part of the functional range at any rate.*
>
> *Gökhan Enç, SAP Deutschland AG & Co. KG*

> *Not just different versions of each single piece of master data must be archived and retrievable, but also versions of other, connected master data constituting an information context or structure.*
>
> *Jan Appl, Mieschke Hofmann & Partner*

[1] The Sarbanes-Oxley Act, for example, specifies in §1520 to store audit records for a period of five years.

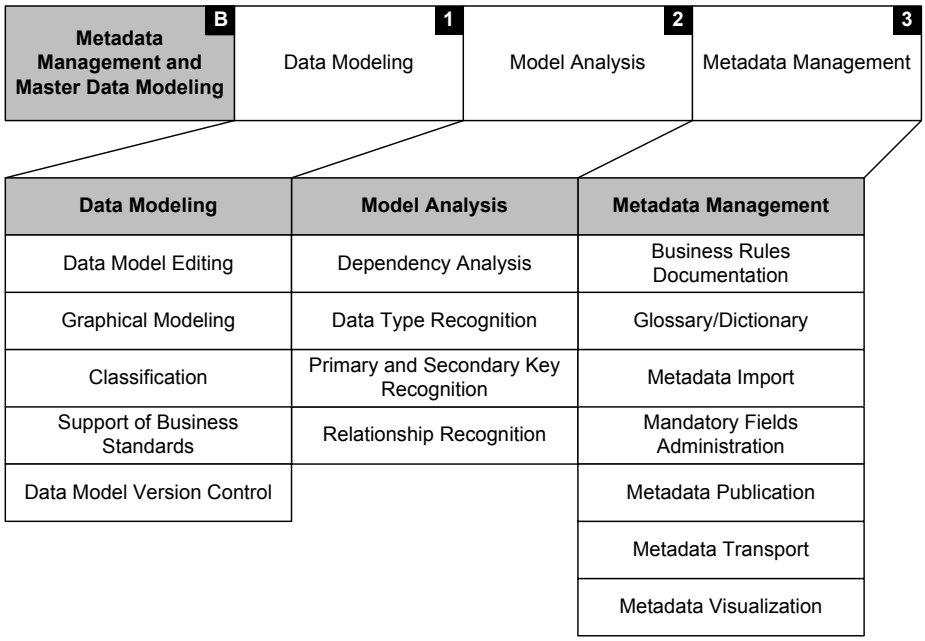


Figure 3-3: Functions of Metadata Management and Master Data Modeling

3.3 Metadata Management and Master Data Modeling

3.3.1 Overview

Basically, metadata describe data characteristics and the meaning of data. In doing so, metadata describe data structures and – in the form of unambiguous specification – ensure correct usage of data throughout an organization [Tozer 1999, Marco 2000]. From an MDM perspective, metadata comprise all the information necessary for efficient management and effective usage of master data. So modeling master data basically means that metadata are created which describe the structure (i.e. type of data, relational cardinalities) of master data. Figure 3-3 shows the functional areas and the functions of this category.

3.3.2 Data Modeling

Data Model Editing

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

By means of *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

Classification allows to group and categorize master data. Assigning data objects to certain categories must not necessarily be unambiguous. Assigning attributes (labeling) allows dynamic classification and dynamic hierarchy building.

Support of Business Standards

Business standards (e.g. eClass, ETIM) facilitate integration of different information systems and create a common understanding of data structures across company or departmental boundaries. *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).

Data Model Version Control

Data Model Version Control allows to archive different versions of any data model. This function ensures that changes in a data model are made transparent and that a data model can be reconstructed as it was at a certain point in time.

> *Apart from being able to archive master data a good product needs to make sure that models and schemes can be archived and version controlled too, if possible automatically.*
>
> *Barbara Bielikova, ZF Friedrichshafen AG*

> *Version control must be part of change management, as there are often transitional phases in which various versions of one and the same data model are used in operating business.*
>
> *Jan Appl, Mieschke Hofmann & Partner*

3.3.3 Model Analysis

Dependency 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

Data Type Recognition allows automatic recognition of data types of existing data objects and automated recognition of data models of existing data sets (e. g. when consolidating two or more different customer data sets).

Primary Key Recognition and Secondary Key Recognition

Primary Key Recognition allows automatic identification of single attributes (or sets of attributes combined) suited to work as the primary key of a certain class of data. *Secondary Key Recognition* checks the key integrity (e.g. the unambiguousness of an attribute when used as (part of) a foreign key).

Relationship Recognition

By means of *Relationship Recognition* relations between types of data are automatically recognized. By this, consolidation of different data inventories is supported.

3.3.4 Metadata Management

Business Rules Documentation

Business rules specify process guidelines or work instructions in the form of If-Then statements. *Business Rules Documentation* supports the communication of (partially formalized) business rules, in order to simplify their use and to keep their definitions up to date. Business rules may refer to strategic decisions (for example, the takeover of another company at a certain share price) or to automated system activities (for example, upon entry of a certain supplier number the maximum quantity to be ordered from this supplier is indicated).

> *Reuse of existing business rules in a universal context really is a very important option an MDM tool should provide.*
>
> *Bernd Binder, Steria Mummert Consulting AG*

> *Documentation of relations between business rules as well as their relations with the master data models and structures should be an essential element of a concept for MDM, since these relations are necessary to understand a business transaction in its entire dimension.*
>
> *Jan Appl, Mieschke Hofmann & Partner*

Glossary/Dictionary

A glossary or dictionary for MDM clearly defines a company's central business objects (e.g. master data such as *customer* or *supplier*) or other elements needed to ensure smooth business processes (e.g. SAP fields).

Metadata Import

Metadata Import allows to consolidate metadata of different formats (available, for example, in text documents or in spreadsheets). This function is important as metadata often are stored in distributed systems and in heterogeneous, partially unstructured formats.

Metadata Transport

Metadata Transport allows automatic transfer of metadata from test systems to transaction systems. This function is important as data structures usually are not created in transaction systems but in special test systems where they can be examined under close-to-reality conditions.

> *No-one would set up and activate a data model directly in an application system without prior testing. So a tool for MDM should provide a metadata transfer functionality allowing to transfer metadata from test systems to application systems.*
>
> *Barbara Bielikova, ZF Friedrichshafen AG*

Mandatory Fields Administration

Mandatory Fields Administration allows central configuration and administration of input fields specified as mandatory fields (e.g. for SAP transactions).

> *In order to ensure quality of data, a function for defining mandatory fields is really important, particularly since data entered into one system often are distributed to subsequent systems.*
>
> *Barbara Bielikova, ZF Friedrichshafen AG*

> *Apart from administration of mandatory fields, what is important too is a clear distinction (based on templates, most preferably) between attributes of objects that are to be maintained centrally, which may not be modified locally then, and attributes of objects that are to be maintained locally, which may not be modified centrally then.*
>
> *Bernd Binder, Steria Mummert Consulting AG*

Metadata Publication

By means of *Metadata Publication* metadata (e.g. business rules or information about business objects) are made available for being used in application systems, where they can be requested with minimum effort (e.g. by placing the cursor above a word).

Metadata Visualization

Metadata Visualization uses metadata (threshold values, business rules templates etc.) in order to graphically display complex phenomena in a simplified manner (e.g. traffic lights graphics for indication of threshold values, or diagrams showing measured values), thereby facilitating data interpretation.

3.4 Master Data Quality Management

3.4.1 Overview

Master Data Quality Management comprises all functions of preventive and reactive master data quality management. Three functional areas can be distinguished: *Data Analysis* provides functions for identifying problems with master data, *Data Aug-*

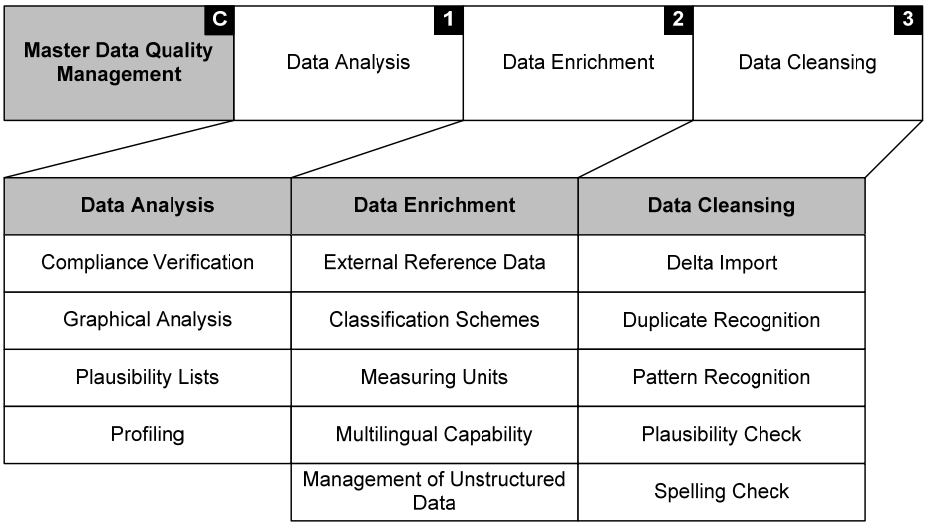


Figure 3-4: Functions of Master Data Quality Management

mentation provides functions for improving master data quality (e.g. by comparison with and integration of external reference data, or by linking images), and *Data Cleansing* provides functions for eliminating errors. Figure 3-4 shows the functional areas and the functions of this category.

> *Data quality is not an end in itself. Assessment of data quality should always take into account how the data are actually used for business, and what the data are used for. Any tool used to measure data quality should be capable of meeting company specific requirements.*
>
> *Ralf Jäger, Client Vela GmbH*

3.4.2 Data Analysis

Compliance Verification

Compliance Verification allows to verify master data against certain guidelines or legal provisions. For example, the Sarbanes-Oxley Act involves a blacklist specifying that doing business with companies from certain countries or with certain companies or persons is prohibited.

> *When sets of master data are created, a compliance engine should be integrated into the process at the earliest stage possible. Verification should refer not just to the buyer but should aim also at invoice recipients, goods recipients, or regulators. Furthermore, a compliance engine should support changes made during the process of order processing, for example when the goods recipient needs to be*

> *changed or when a new customer account needs to be created. If it is detected that compliance is not given, both the master data set and the respective business process need to be blocked. Upon clearance of the matter, both can be approved again by special authorities only. Last but not least, it must be made sure that compliance verification uses up-to-date verification data.*
>
> *Karlheinz Sturm, Voith Turbo GmbH & Co.KG, Heidenheim*

Graphical Analysis

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

Plausibility Lists provides a basis for other functions (*Profiling*, *Plausibility Check*). Plausibility lists may contain reference data (e.g. company addresses), domains of definition for numeric master data attributes (e.g. dimensions, weight), or regular expressions specifying the structure of, for example, DUNS numbers, phone numbers, e-mail addresses, or dates.

> *It would be good to have free access to databases of registration authorities in order to be able to verify general information on companies, such as name, company address etc. Currently, such services are unsatisfying and expensive. This is a problem practically each company has.*
>
> *Helge Enenkel, Voith Paper Holding GmbH & Co. KG*

Profiling

Profiling allows to analyze data and, based on predefined rules (e.g. *a name must not contain a ‚?'*), create a statistical profile regarding compliance with such rules. This profile, in turn, provides a basis for other functions (e.g. *Duplicate Record Recognition*).

> *I am currently not aware of any other MDM tool providing sufficient profiling functionality. I consider such functionality important in order to be able to get a quick understanding of the data's constitution.*
>
> *Manfred Nielsen, KARL STORZ GmbH & Co. KG*

3.4.3 Data Enrichment

External Reference Data

External Reference Data allows to substitute missing data by external data (e.g. the company address register provided by a telecommunications company) or to match existing data with external data in order to detect errors in one's database.

> *Country specific differences must be taken into consideration particularly with regard to customer master data. The way company addresses are structured, for example, is highly different from one country to another, even within Europe. Thus, it is quite difficult to represent company addresses in a uniform data structure.*
>
> *Ralf Jäger, Client Vela GmbH*

Classification Schemes

Classification Schemes supports the use of classification systems (for example eClass, ETIM) for corporate MDM.

> *The linking of master data repositories with classification schemes such as eClass, EDMA or UNSPSC should really be taken in to account by MDM, particularly for being able to create catalogs and do good reporting.*
>
> *Jörn Bachmeier, Roche Diagnostics GmbH*

Measuring Units

Measuring Units supports conversion of measuring units (e.g. attributes of dimension or weight). Particularly companies acting on a global level are confronted with various measuring units, for example in customs or business-to-business scenarios.

Multilingual Capability

Multilingual Capability allows to make master data and metadata available in various languages at constant data consistency. Making master data available to users in several languages may be necessary to increase data interpretability and process performance, thereby improving data quality.

> *Multilingual capability should comprise intelligent search functionality when using special, national characters. Data consistency must be ensured in several respects, such as a) for fields that are independent of language and that have the same content in every language (e.g. zip codes), b) for key fields, which are internally identical but have different meanings for external users (e.g. country codes), and c) for fields that have different content in different languages (e.g. names of cities, such as München / Munich).*
>
> *Helge Enenkel, Voith Paper Holding GmbH & Co. KG*

Management of Unstructured Data

Management of Unstructured Data allows efficient administration of unstructured data (CAD drawings, photos, videos, digitalized records etc.) and their relations with master data objects, as well as efficient provision of such data, which is particularly required in manufacturing processes, but also in marketing processes.

3.4.4 Data Cleansing

Delta Import

See *Delta Import* (*3.5.2 Data Import*). Identification of the Delta (i.e. of what was changed) can be useful to search for duplicate records, for example.

Duplicate Recognition

Duplicate Recognition allows to search for duplicate records. Also, this function generates warnings during data entry indicating data duplication. Automatic identification of duplicate records is a complicated matter for which there can be no generic solution. Whether duplicate records exist depends on the context data is used in. So duplicate record recognition must be based on specific rules, for which this function provides templates (e.g. *minimal editing distance* for text entries, or v*ariance* for numerical values).

Pattern Recognition

Pattern Recognition identifies certain patterns in data repositories. Patterns (generated, for example, by regular expressions) allow to define expected data structures (e.g. of a phone number, an e-mail address, or a DUNS number) or invalid entries

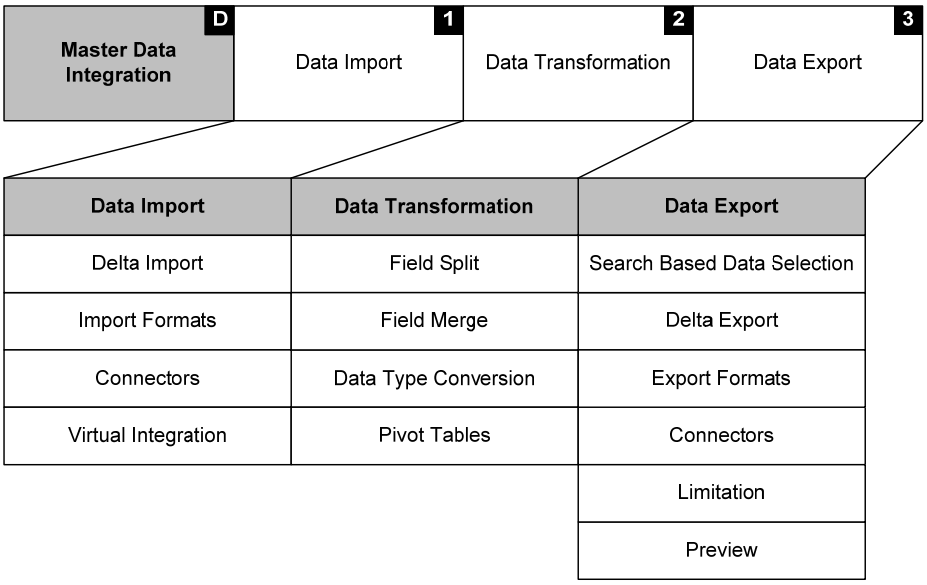


Figure 3-5: Functions of Master Data Integration

(e.g. a zip code that is too long, or a missing street number). If errors occur, the function provides defined responses (e.g. generation of a warning, or automatic correction of an invalid value).

Plausibility Check

See *Plausibility Check* (*3.2.2 Data Creation*).

Spelling Check

Spelling Check corrects typical mistakes occurring during data entry, such as names or terms written in a wrong way. The function can be supported by reference lists used also for *Plausibility Check* (*3.2.2 Data Creation*).

3.5 Master Data Integration

3.5.1 Overview

Master Data Integration comprises functions supporting transfer (import and export) and structural transformation (e.g. consolidation of fields or tables) of master data. Figure 3-5 shows the functional areas and the functions of this category.

> *When MDM systems are to be implemented, business realities should be taken into consideration. By intelligently integrating legacy systems into a new MDM solution it is possible to accomplish smooth transition, allowing effective control of*

> *the implementation process and integration of valuable knowledge. A good MDM system should support this process by offering flexible and easily adaptable interfaces.*
>
> *Ralf Jäger, Client Vela GmbH*

3.5.2 Data Import

Delta Import

Delta Import allows to import data created or modified since the previous import (the Delta).

Import Formats

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. The function offers support for importing XML based data structures, spreadsheets, or structured flat files.

Connectors

Connectors allows to create new interfaces for, for example, importing data of a format originally not supported. Such connectors usually are offered as transformation languages (e.g. XSLT) or APIs.

Virtual Integration

Virtual Integration allows to temporarily bring together data from different source systems without needing to copy them into a common database. Transformed data are directly transferred to the source systems, i.e. they are not saved in a central system from which they need to be exported at a later stage.

3.5.3 Data Transformation

Field Split

Field Split allows to split values of one field into several values, following predefined rules (e.g. by a separator , _ ‘, or , ; ‘).

Field Merge

Field Merge allows to merge values of several fields.

Data Type Conversion

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

Pivot Tables allows to restructure data classes structured in tables (e.g. by new ordering schemes or inserting rows and columns).

3.5.4 Data Export

Search Based Data Selection

Search Based Data Selection allows the explicit selection of data objects to be exported from a list (result of a search query).

Delta Export

Cp. *Delta Import* (*3.5.2 Data Import*).

Export Formats

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

See *Connectors* (*3.5.2 Data Import*)

Limitation

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

Preview allows to view data to be exported as they will be provided.

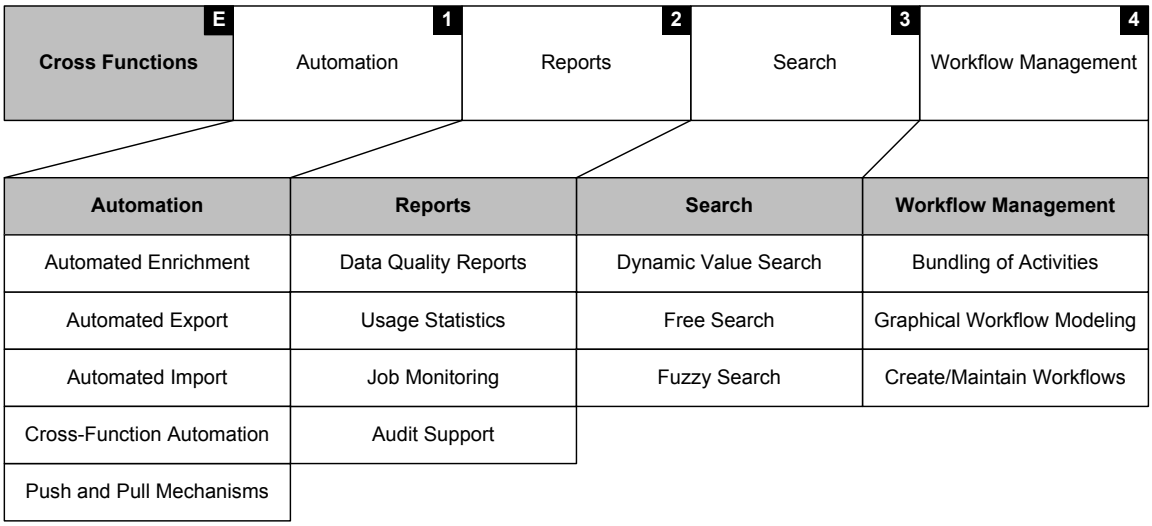


Figure 3-6: Cross Functions

3.6 Cross Functions

3.6.1 Overview

Functions that cannot be assigned to one of the previous categories are subsumed under the *Cross Functions* category. The functions under the functional area *Automation* do not provide additional functionality but offer support for being able to efficiently use other functions from other functional areas. Some functions under the functional area *Automation* take up results generated by functions under the functional area Data Analysis and make them ready for further processing by machines or humans. Figure 3-6 shows the functional areas and the functions of this category.

3.6.2 Automation

Automated Enrichment

Cp. *3.4.3 Data Enrichment*.

Automated Export

Cp. *3.5.4 Data Export*. *Automated Export*, together with *Automated Import*, allows to build a system for automated exchange of master data between a test system and transaction systems.

Automated Import

Cp. *3.5.2 Data Import*. *Automated Import*, together with *Automated Export*, allows to build a system for automated exchange of master data between a test system and transaction systems.

> *When it comes to automation, monitoring is something that is extremely critical in order to ensure smooth and robust operation. Particularly if local systems cover several time zones, tracing data exports and imports can quickly become difficult if there are no supporting tools at hand.*
>
> *Bernd Binder, Steria Mummert Consulting AG*

Cross-Function Automation

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

Push and Pull Mechanisms allows to apply both push-mechanisms and pull-mechanisms for automated data import and export. Automated data import frequently is done upon a push-mechanism (e.g. periodically triggered by a central master data server). Other data import scenarios, however, may require a push- mechanism (e.g. if a change in data is made in a local system). For automated data export, the same principle applies vice versa.

> *As the master data management system is not able to recognize whether data have been updated in local systems, a push-mechanism should be provided too.*
>
> *Gökhan Enç, SAP Deutschland AG & Co. KG*

3.6.3 Reports

Data Quality Reports

Data Quality Reports allows to illustrate the results of data analyses (see *3.4.2 Data Analysis*), e.g. by diagrams to be used in dashboards, or by preconfigured templates for management reports.

Usage Statistics

Usage Statistics allows to record in real-time who is using or requesting which data. The function also provides a ranking list, which can be used, for example, for system development, in terms of making sure that important data is available at any time.

Job Monitoring

Job Monitoring allows to monitor automated functions (see *3.6.2 Automation*) and assess them by various indicators (e.g. processing time, error rate).

> *It is important to define KPIs and other target values from the outset against which the system can be tested and evaluated during regular operation.*
>
> *Bernd Binder, Steria Mummert Consulting AG*

Audit Support

Audit Support helps create (e.g. by providing templates or preconfigured analyses, see *3.4.2 Data Analysis*) reports demanded by legal provisions[1].

3.6.4 Search

Dynamic Value Search

Dynamic Value Search allows to search for and identify data objects by means of known attribute values. The function is supported by dynamic sorting and filtering mechanisms that can be applied to single attributes (e.g. when a certain sequence of letters is entered, all names are listed that begin with this sequence of letters, with the list dynamically changing every time another letter is entered).

1 Article 33 of REACH (Registration, Evaluation, Authorization, and Restriction of Chemicals), a regulation by the European Union, specifies that companies are obliged to notify the European Chemicals Agency of 'chemical substances of very high concern' if such a substance is present at more than 0.1% of the mass of a product. A respective report must be provided within 45 days.

Free Search

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

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*). *Fuzzy Search* is an important function particularly when searching across multilingual databases, in order to facilitate, for example, the search for words containing language specific, special characters (e.g. *Joel/Joël*, or *München/Munchen/Muenchen*).

3.6.5 Workflow Management

Bundling of Activities

Bundling of Activities allows to bundle several activities within a single workflow. This function is important as MDM workflows may comprise a number of detailed instructions (e.g. verification of single attribute values, or creation of several customer data objects).

> *Apart from viewing MDM from a perspective focusing on technical functionality and business requirements, companies should take into account also organizational aspects, such as consistent process definitions. A good workflow engine as part of an MDM system could bring about more freedom regarding modeling of processes.*
>
> *Ralf Jäger, Client Vela GmbH*

Graphical Workflow Modeling

Graphical Workflow Modeling allows to model workflows by means of graphical symbols (e.g. rectangles for activities, arrows for modeling a sequence of activities).

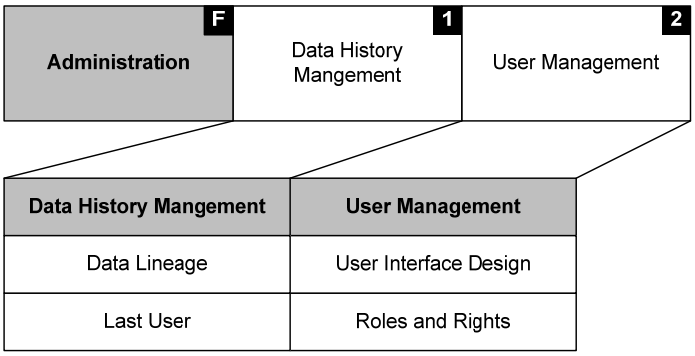


Figure 3-7: Functions of Administration

Create/Maintain Workflows

Create/Maintain Workflows allows to manage sequences of activities across processes and departments. This function is important as, along the entire data lifecycle, numerous activities are executed by numerous people, each time modifying a data object's state (see Section 3.2).

> *Using a special workflow management tool for MDM does not seem to make too much sense. Although workflow management is an indispensable element of MDM, it should be done with existing tools.*
>
> *Hans Jakob Reuter, gicom GmbH*

> *As concrete requests to create or alter master data typically come from local systems and not from a central system, the respective workflow must comprise the entire process, without any media breakage, from local expression of a request to the processing of the data in a central system.*
>
> *Bernd Binder, Steria Mummert Consulting AG*

3.7 Administration

3.7.1 Overview

The category *Administration* comprises functions for user administration and change management. Figure 3-7 shows the functional areas and the functions of this category.

3.7.2 Data History Management

Data Lineage

Data Lineage allows to trace back the origin of pieces or sets of master data. This function is particularly important if master data from various, distributed information systems were consolidated in a central MDM server. Using *Data Lineage*, the source system of any attribute value of any data object can be identified at any time.

Last User

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.

> *An MDM system should provide a detailed data modification history, with any modification made being documented by information such as content of entry field now and before, date of data modification, and user ID. Data modification history must inform about the 'Who-What-When', not about the 'Why'. It should also be possible to post comments with or attach documents to the data object (for example, a letter from a business customer announcing a change in the legal form of the company as of a certain date).*
>
> *Helge Enenkel, Voith Paper Holding GmbH & Co. KG*

3.7.3 User Management

User Interface Design

User Interface Design allows to adapt the graphical user interface to meet role specific requirements. The graphical user interface should provide only functions needed by a certain role or which a certain role is entitled to use. Also, it should be possible to design functions (and/or the results they generate) in such a way that roles are best supported in their specific processes.

> *The graphical user interface of an MDM tool has an enormous influence on data quality as a whole.*
>
> *Gökhan Enç, SAP Deutschland AG & Co. KG*

Roles and Rights

Roles and Rights allows to define roles and to assign entitlements to execute certain activities by such roles.

4 Product Analysis

4.1 Introduction

The following paragraphs contain descriptions of products for MDM offered by IBM, Oracle, SAP, and TIBCO. As each product presented can usually be part of an overall solution, which must be adapted to company specific requirements, the products are described only roughly. To get more detailed information we recommend to check the documentation provided by the manufacturers themselves.

As the present paper primarily aims at establishing a common terminology for MDM functions, the selection of products should be considered exemplary, with the product descriptions by no means being exhaustive.

All products examined consist of components providing functions the use of which would require an integration platform (usually software implementing a service oriented architecture (SOA)). So the descriptions of the products follow the structure components – functions – integration platform. Section 4.6 offers an overview as to whether and how each product covers the range of functions specified in the Functional Reference Architecture.

4.2 IBM InfoSphere

IBM InfoSphere is a product family providing a platform capable of integrating, matching, managing, and analyzing data from systems of different manufacturers, and making available such data to users, application systems, and business processes. Components of the product family are *IBM InfoSphere Warehouse*, *IBM InfoSphere Information Server*, and *IBM InfoSphere MDM Server*, the services of which can be flexibly combined over a SOA.

IBM InfoSphere Information Server supports companies in using and managing distributed data. This software platform consists of several modules providing analysis, cleansing, and integration of data coming from disparate sources, with basic services

(such as meta data management, access management, or parallel processing) being provided by each module. Modules of *InfoSphere Information Server* are:

- *InfoSphere DataStage*. Offers integration functions for processing of both structured and unstructured data coming from disparate data repositories, ERP systems, or other transaction systems.
- *InfoSphere QualityStage*. Offers functions for matching and standardizing data coming from disparate sources.
- *InfoSphere Information Analyzer*. Supports creation of source system profiles (analysis of interfaces) and monitors compliance with rules supposed to prevent usage and dissemination of poor or invalid data.
- *InfoSphere Metadata Server*. Offers various functions for analyzing, modeling, using, and managing metadata (data structures and descriptions of business objects) based on a meta-model that can be adapted to company specific requirements.
- *InfoSphere Information Services Director*. Allows publication of data access and integration processes as reusable services in a SOA.
- *Connectivity Software*. Offers cross functions and administration functions, such as functions specifying rights and roles (*Security Services*), functions for generating reports, functions for tracing back the origin of data (*Log Services*), or functions allowing realtime access to data sources (e.g. by means of *InfoSphere Federation Server*, *InfoSphere Replication Server,* or *InfoSphere Change Data Capture*).

IBM InfoSphere MDM Server allows centralized MDM. It enhances the functionality of *IBM InfoSphere Information Server* by MDM specific services (such as identification of important customers, management of relationships between master data objects, lunching of new products and product packages, publication management), which are also made available in a SOA.

4.3 Oracle Master Data Management Suite

Oracle Master Data Management Suite comprises several products which can be combined using a SOA named *Fusion Middleware*. Components of *Oracle Master Data Management Suite* are:

- *Business Rules*. Allows graphical modeling, automated execution, and monitoring of business rules. Such business rules may, for example, specify definitions of duplicate records or monitor workflows.
- *Customer Data Hub* / *Product Data Hub*. Provide standardized models for customer / product data and various lifecycle functions for creating, maintaining, and archiving customer / product data. Both components comprise also preconfigured rules for identifying duplicate records and for plausibility checks, and they offer interfaces to external data providers.
- *Data Integrator*. Provides functions for statistical data analysis, based on rules (similar to *Business Rules*). The rules are also used in the evaluation process as metrics.
- *Data Quality Matching Server* / *Data Quality Cleansing Server*. Provides functions for searching and identifying duplicate records / for automated elimination of identified errors.
- *Hyperion Data Relationship Management*. Provides functions for definition and management of data structures and classification schemes, and supports audits through recording and archiving of changes.

4.4 SAP NetWeaver Master Data Management

SAP NetWeaver Master Data Management (SAP NetWeaver MDM) is part of the *SAP NetWeaver* technology platform. It comprises far-reaching integration functions for master data exchange, portal integration, business warehousing, monitoring, transportation etc. Apart from the server instances, *SAP NetWeaver MDM* mainly consists of four client components for management, editing and processing, and import and

export of master data. MDM functionality is complemented by components from the SAP product portfolio (e.g. *SAP BusinessObjects Data Services*).

SAP NetWeaver MDM operates on an own database that works as a buffer (at least during the time of data transformation). The database may also serve as a central master data repository making master data available to transaction systems. For integration purposes, *SAP NetWeaver MDM* can be embedded in *SAP Process Integration (SAP PI)*, which is a component that supports data exchange across processes and application systems. Components of *SAP NetWeaver MDM* are:

- *MDM Console*. Provides functions for data modeling and user management. Data structures are created in the form of tables (or relations between such tables) and are stored in repositories. These repositories are used by other components for instantiating the structures with data (*SAP Import Manager*) or for modifying data (*SAP Data Manager*).
- *MDM Import Manager*. Allows to process data of various formats (e.g. over interfaces with databases, or from spreadsheets) and to import such data into the repositories defined by *MDM Console*.
- *MDM Data Manager*. Provides functions for data transformation and serves as a central tool for ad-hoc analyses, amendments, corrections, or searches. For adaptation of the functionality to user specific requirements, functions can be made available over a Web based portal.
- *MDM Syndicator*. Allows to export data from repositories defined by *MDM Console*. Automated export is possible using *MDM Syndication Server*.

MDM functionality is further complemented by *SAP Business Objects Data Services*, providing further tools for data quality analysis and services for data quality improvement (e.g. cleansing and validation of address data by means of reference data). Furthermore, *SAP Business Objects Metadata Management* offers functions for veri-

fication and consolidation of metadata, and functions for putting up a glossary for business objects.

4.5 TIBCO Collaborative Information Manager

Collaborative Information Manager (CIM) is TIBCO's solution for MDM. The functions of *CIM* are available both over a Web based interface and as Web Services, and can be used and configured with various integration platform (e.g. *WebSphere* by IBM, *ActiveMatrix* by TIBCO). Modeling of MDM related workflows is possible over a Web based interface by means of a graphical modeling tool, which is based on *Eclipse*, an integrated development platform. *CIM-Engine* allows to manage master data and the relations between them.

CIM comprises the following components, the functionality of which can be used very flexibly using combined Web Services (Composite Services):

- *Information Repository*. Contains the data structures and offers various metadata management functions for data modeling, data classification, and model management. Also, the component comprises Web Services for data search and data versioning.
- *Data Governance*. Provides functions for definition, execution, and monitoring of business rules. These rules are implemented by the so-called *Complex Event Processing Engine*, which can also be used for roles and rights management.
- *Process Automation*. Provides functions for definition, administration, execution, and monitoring of workflows.
- *Synchronization*. Comprises functions for both manual and automated data import and export, with industry standards for business-to-business data exchange partially being supported.
- *Reporting* and *Business Intelligence*. Provide functions for reporting, measuring, and data quality improvement, as well as for optimization of the performance of MDM workflows administered by *Process Automation*.

4.6 Coverage of Functional Reference Architecture

Table 4-1 shows how each of the four products examined covers the range of functions specified by the Functional Reference Architecture.

Function	IBM	Oracle	SAP	TIBCO
Master Data Lifecycle Management				
Data Creation				
Conditional Entries	InfoSphere MDM Server		MDM Data Manager	Information Repository (conditional relations)
Bulk Editing	InfoSphere MDM Server		MDM Data Manager	Information Repository
Plausibility Check	InfoSphere MDM Server	Data Hubs	MDM Data Manager	Information Repository (Validation Rules)
Data Maintenance				
Check-out			MDM Data Manager	Information Repository
Bulk Editing	InfoSphere MDM Server		MDM Data Manager	Information Repository (via DBLoader)
Plausibility Check	InfoSphere MDM Server	Data Hubs	MDM Data Manager	Information Repository (Validation Rules)
Data Deactivation				
Bulk Editing	InfoSphere MDM Server	Data Hubs	MDM Data Manager	Information Repository
Data Archiving				
Archiving	IBM Optim	Hyperion DRM	MDM Console	Information Repository
History Control	InfoSphere MDM Server	Hyperion DRM	MDM Console	Information Repository
Metadata Management and Master Data Modeling				
Data Modeling				
Data Model Editing	InfoSphere Data Architect	Hyperion DRM	MDM Console	Information Repository
Graphical Modeling	InfoSphere Data Architect			Information Repository
Classification	InfoSphere Business Glossary	Hyperion DRM	MDM Data Manager	Information Repository

Function	IBM	Oracle	SAP	TIBCO
Support of Business Standards		Data Hubs	MDM Console	Synchronization
Data Model Version Control	InfoSphere Data Architect	Hyperion DRM	SAP NetWeaver (Change and Transport System)	Information Repository
Model Analysis				
Dependency Analysis	InfoSphere Metadata Workbench		MDM Import Manager	Information Repository
Data Type Recognition	InfoSphere Information Analyzer	Data Integrator	MDM Import Manager	Synchronization
Primary Key Recognition and Secondary Key Recognition	InfoSphere Information Analyzer		MDM Import Manager	Information Repository
Relationship Recognition	InfoSphere Information Analyzer	Data Integrator	MDM Import Manager	Synchronization
Metadata Management				
Business Rules Documentation	ILOG	Business Rules	SAP NetWeaver (Business Process Management)	Data Governance (via RuleBases)
Glossary/Dictionary	InfoSphere Business Glossary		SAP BO Metadata Management	Information Repository (via work flow)
Metadata Import	InfoSphere Metadata Workbench		SAP BO Metadata Management	Synchronization
Metadata Transport	InfoSphere Metadata Workbench		SAP NetWeaver (Change and Transport System)	
Mandatory Fields Administration	InfoSphere Metadata Workbench	Business Rules		Information Repository (Validation Rules)
Metadata Publication	InfoSphere Metadata Workbench		SAP BO Metadata Management	
Metadata Visualization	InfoSphere Metadata Workbench			

Function	IBM	Oracle	SAP	TIBCO
Master Data Quality Management				
Data Analysis				
Compliance Verification	InfoSphere Information Analyzer	Business Rules	SAP BO BusinessObjects Data Services	Data Governance (Business Rules)
Graphical Analysis	InfoSphere MDM Server		SAP BO BusinessObjects Data Services	
Plausibility Lists	InfoSphere Quality Stage	Data Integrator	SAP BO BusinessObjects Data Services	Reporting und BI
Profiling	InfoSphere Quality Stage	Data Hubs	SAP BO BusinessObjects Data Services	
Data Enrichment				
External Reference Data	InfoSphere Quality Stage	Data Hubs	SAP BO BusinessObjects Data Services	
Classification Schemes	InfoSphere Quality Stage	Data Hubs	MDM Console	Information Repository
Measuring Units	InfoSphere Quality Stage		MDM Data Manager	Information Repository
Multilingual Capability	InfoSphere Business Glossary		MDM Data Manager	Information Repository
Management of Unstructured Data	InfoSphere MDM Server, IBM Content Manager		MDM Data Manager	Information Repository
Data Cleansing				
Delta Import	InfoSphere MDM Server			Synchronization
Duplicate Recognition	InfoSphere QualityStage, InfoSphere MDM Server	Data Quality	MDM Import Manager	Data Governance
Pattern Recognition	Quality Stage	Business Rules	Expressions (in MDM Data Manager)	Data Governance
Plausibility Check	InfoSphere MDM Server	Data Hubs		Data Governance
Spelling Check	InfoSphere Qualiy Stage			

Function	IBM	Oracle	SAP	TIBCO
Master Data Integration				
Data Import				
Delta Import	InfoSphere MDM Server		MDM Import-Server, SAP BO Data Services	Synchronization
Import Formats	InfoSphere Data Stage	Data Integrator	MDM Import-Server, SAP BO Data Services	Synchronization
Connectors	InfoSphere Data Stage	Application Integration	Process Integration, SAP BO Data Services	Synchronization
Virtual Integration	InfoSphere Federation Server	Data Integrator		Synchronization
Data Transformation				
Field Split	InfoSphere Qualiy Stage	Data Quality	MDM Import Manager	Data Governance
Field Merge	InfoSphere Data Stage	Data Quality	MDM Import Manager	Data Governance
Data Type Conversion	InfoSphere Data Stage	Data Integrator	MDM Import Manager	Data Governance
Pivot Tables	InfoSphere Data Stage		Pivot	
Data Export				
Search Based Data Selection	InfoSphere MDM Server		MDM Syndicator	Information Repository
Delta Export	InfoSphere MDM Server		MDM Syndicator	Synchronization
Export Formats	InfoSphere Data Stage	Application Integration	MDM Syndicator	Information Repository / Synchronization
Connectors	InfoSphere Data Stage	Application Integration	API / Web services	Information Repository / Synchronization
Limitation			MDM Syndicator	Synchronization
Preview	InfoSphere Data Stage		MDM Syndicator	Synchronization (via Validate Synchronization Profile)

Function	IBM	Oracle	SAP	TIBCO
Cross Functions				
Automation				
Automated Enrichment	InfoSphere MDM Server	Data Quality		Information Repository (via Defined Data Source)
Automated Export	InfoSphere MDM Server		MDM Syndication Server	Synchronization
Automated Import	InfoSphere MDM Server		MDM Import Server	Synchronization
Cross-Function Automation	InfoSphere Information Server			All components (via Web services)
Push and Pull Mechanisms	InfoSphere MDM Server			Synchronization
Reports				
Data Quality Reports	InfoSphere MDM Server	Data Integrator	SAP BO Data Services	Reporting und BI
Usage Statistics	InfoSphere MDM Server			
Job Monitoring	Log Services, InfoSphere MDM Server		SAP BO Data Services	Reporting und BI
Audit Support	InfoSphere MDM Server			
Search				
Dynamic Value Search	InfoSphere MDM Server		MDM Data Manager	
Free Search	IBM Omnifind		Portal	Information Repository
Fuzzy Search	InfoSphere MDM Server		SAP BO Data Services	Information Repository
Workflow Management				
Bundling of Activities	WebSphere Process Server	Business Rules		Process Automation
Graphical Workflow Modeling	WebSphere Process Server		MDM Data Manager (Integration von Microsoft Visio)	Process Automation
Create/Maintain Workflows	WebSphere Process Server	Business Rules	MDM Data Manager (Integration von Microsoft Visio)	Process Automation

Function	IBM	Oracle	SAP	TIBCO
Administration				
Data History Management				
Data Lineage	InfoSphere MDM Server	Hyperion DRM		Information Repository (via custom attributes)
Last User	InfoSphere MDM Server		User Stamp (via MDM Console)	Reporting und BI
User Management				
User Interface Design	WebSphere Portal		MDM Data Manager / Portal	Data Governance
Roles and Rights	InfoSphere MDM Server		MDM Console	Data Governance

Table 4-1: Coverage of Functional Reference Architecture

5 Selected Case Studies

5.1 SAP NetWeaver MDM at Oerlikon Textile

Oerlikon Textile GmbH & Co. KG, with headquarters in Switzerland, is a manufacturer of textile machinery and a provider of textile industry solutions covering the entire value chain. Employing more than 7,400 people, the company has branches for development, manufacture, and distribution in more than fifty locations worldwide (covering, for example, the North American market, the Middle East market, and the emerging markets of China and India).

Oerlikon's business units are supported by seven independent systems based on the SAP ERP application. Each unit keeps and maintains its own business partner master data. Over time, more and more data silos came into existence, leading to more and more isolated data repositories partially containing conflicting data. Inconsistent and redundant master data led to problems particularly when it came to coordinating business activities involving two or more business segments of the company, which to some extent operate in the same markets and address the same business partners. So, in order to be successful in the long run, what was needed was a common, harmonized database.

> *We consider a common database as indispensable for being able to effectively coordinate market activities across business units in order to exploit business opportunities as fully as possible.*
>
> *Claudia Siebertz, Project Manager*

In order to achieve such master data harmonization, Oerlikon Textile searched for a solution capable of consolidating data from disparate source systems and making harmonized data available to all users working in the seven business units. The German business segment of Oerlikon Textile, which is located in Remscheid, took over coordination of the global master data harmonization project. It was decided to

use *SAP NetWeaver MDM*, as this product was considered to suit perfectly the needs of Oerlikon Textile regarding master data consolidation and harmonization.

> *With the centralized management of all customer-related data, we decided on a very broad deployment.*
>
> *Claudia Siebertz, Project Manager*

From several possible harmonization concepts Oerlikon Textile chose the one aiming at centralizing the management of business partner master data. To do so, it was first necessary to consolidate, cleanse, and update all data repositories accommodated by the SAP ERP systems. This effort took six months, with SAP Consulting offering support during implementation. Estimates put up by SAP experts prior to the project's start turned out to be stable, so that deadlines were met and costs remained within an acceptable frame.

> *Benefiting from specific product and industry know-how, from transfer of knowledge from other projects, and from the commitment of the consultants – those were the points reassuring us that working with SAP Consulting was the right decision.*
>
> *Claudia Siebertz, Project Manager*

Today all business partner master data are being managed and maintained centrally at Oerlikon Textile by a special team comprising people from all business units. Effective MDM is ensured by the team's proximity to customers and markets. Specific workflows and automation support the change and approval processes that relate to managing the data. After having been updated and harmonized master data then are made available to target systems all over the world by means of interactive distribution mechanisms.

The individual business units of Oerlikon Textile now benefit from high consistency and relevance of master data facilitating cross-unit business activities and helping exploit cross-selling and up-selling potentials. By centralizing MDM, risks and

shortcomings in data processing could clearly be reduced, and the quality of the database could substantially be improved.

5.2 IBM Master Data Management at PostFinance

One of the major challenges in the operating business of banks is to manage customer accounts centrally and in a standardized way. There are various requirements that need to be met. For example, in case of customers who have more than one account, master data must be highly consistent and easily accessible for every user in the company. And to avoid the risks of money laundering and other criminal acts, customer master data must be highly transparent and easy to oversee.

Employing over 3,500 people, PostFinance, a business unit of Swiss Post, is a major provider of financial services in Switzerland. For private customers, the company offers services in the fields of payments, investments, financing, and retirement planning. And for business customers, PostFinance mainly offers innovative payment transaction services and a range of investment models and financing products. Apart from its actual workforce, about 12,000 more people work for PostFinance on a 30 percent to 50 percent level in the 2,500 post offices across Switzerland. In 2008, about 120,000 new customers opted for the services of PostFinance, increasing the number of accounts by 311,000.

> *We are using Linux, Windows, and Unix operating systems, but no hosted systems, what is rather untypical in the banking industry. We searched for an MDM solution that fits into this technical environment.*
>
> *Jochen Schneider, Head of IT Department*

PostFinance has a modern IT infrastructure that is mainly based on a client-server architecture. Master data used to be processed in more than thirty application systems, both by online access and by data replication. Because of this, customer account management became more and more complex and difficult to oversee, particularly when customers had more than one account. The problem was considered as

quite serious also since such limited transparency and usability of customer master data can foster criminal behavior of customers, such as money laundering.

> *The problem of money laundering was one of the driving forces for us to think about a new MDM solution. First we tried to build a new solution with our old systems, but it turned out that this would not take us to where we wanted to be.*
>
> *Jochen Schneider, Head of IT Department*

After a short phase of evaluation PostFinance decided in April 2006 to implement *IBM WebSphere Customer Center* (which is a component of *IBM InfoSphere MDM Server* since the beginning of 2009). Since November 2008, more than 1,000 employees of PostFinance have been working with this new system. Although it has been only a couple of months since the new era of MDM started at PostFinance, the benefits have become quite obvious already.

> *The new solution allows us to establish a fully consistent master data structure, what makes customer management so much easier. Also, our sales department benefits from better, more precise data on customers. And finally, we are now able to effectively tackle the problem of money laundering by using, for example, blacklists.*
>
> *Jochen Schneider, Head of IT Department*

Apart from the benefits for the company itself, the new MDM concept at PostFinance has positive consequences for the customers too.

> *Thanks to our new MDM solution we are now able to open a new account within six minutes. Our customers appreciate that very much.*
>
> *Jochen Schneider, Head of IT Department*

6 Summary and Outlook

The paper presented a Functional Reference Architecture for MDM describing functional requirements software products for MDM should meet from a business perspective. The Functional Reference Architecture was developed with the help of 34 subject matter experts discussing in three focus groups. MDM solutions of four software providers were examined with regard to their capability to meet the functions specified in the Functional Reference Architecture. Future research on the topic should aim at

- evaluating and adapting the structure and the content of the Functional Reference Architecture by more focus group settings, and
- assessing and evaluating more MDM software products against the Functional Reference Architecture.

Literature

[Back et al. 2007]
Back, A., von Krogh, G., Enkel, E., The CC Model as Organizational Design Striving to Combine Relevance and Rigor, in: Systemic Practice and Action Research, 20, 2007, No. 1, pp. 91-103

[Boisot/Canals 2004]
Boisot, M., Canals, A., Data, information and knowledge: have we got it right?, in: Journal of Evolutionary Economics, 14, 2004, No. 1, pp. 43-67

[Bourdreau/Couillard 1999]
Bourdreau, A., Couillard, G., Systems Integration and Knowledge Management, in: Information Systems Management, 16, 1999, No. 4, pp. 24-32

[DAMA 2008]
DAMA, The DAMA Dictionary of Data Management, Technics Publications, 2008

[Davenport/Prusak 1998]
Davenport, T. H., Prusak, L., Working Knowledge: How Organizations Manage What They Know, HBS Press, Boston 1998

[Dreibelbis et al. 2008]
Dreibelbis, A., Hechler, E., Milman, I., Oberhofer, M., van Run, P., Wolfson, D., Enterprise Master Data Management: An SOA Approach to Managing Core Information, IBM Press, Upper Saddle River 2008

[Eickhoff 1999]
Eickhoff, B., Gleichstellung von Frauen und Männern in der Sprache, in: Sprachspiegel, 55, 1999, No. 1, pp. 2-6

[English 1999]
English, L. P., Improving Data Warehouse and Business Information Quality, Wiley, New York 1999

[Heilig et al. 2007]
Heilig, L., Karch, S., Böttcher, O., Hofmann, C., Pfennig, R., SAP NetWeaver Master Data Management, Galileo, Bonn 2007

[Kagermann/Österle 2006]
Kagermann, H., Österle, H., Geschäftsmodelle 2010: Wie CEOs Unternehmen transformieren, Frankfurter Allgemeine Buch, Frankfurt a. M. 2006

[Lee et al. 2006]
Lee, Y. W., Pipino, L. L., Funk, J. D., Wang, R. Y., Journey to Data Quality, MIT Press, Boston 2006

[Marco 2000]
Marco, D., Building and Managing the Meta Data Repository: A Full Lifecycle Guide, Wiley, New York 2000

[Mertens et al. 2004]
Mertens, P., Bodendorf, F., König, W., Picot, A., Schumann, M., Hess, T., Grundzüge der Wirtschaftsinformatik, Springer, Berlin 2004
[Mosley 2008]
Mosley, M., DAMA-DMBOK Functional Framework. Version 3.02, DAMA International, 2008
[Österle/Winter 2003]
Österle, H., Winter, R., Business Engineering, in: Österle, H., Winter, R. (Eds.), Business Engineering - Auf dem Weg zum Unternehmen des Informationszeitalters, Springer, Berlin 2003, pp. 4-20
[Otto/Österle 2009]
Otto, B., Österle, H., Eine Methode zur Konsortialforschung, BE HSG / CC CDQ / 6, Institut für Wirtschaftsinformatik, Universität St. Gallen, 2009
[Pipino et al. 2002]
Pipino, L. L., Lee, Y. W., Wang, R. Y., Data Quality Assessment, in: Communications of the ACM, 45, 2002, No. 4, pp. 211-218
[Redman 1996]
Redman, T. C., Data Quality for the Information Age, Artech House, Boston 1996
[Spiegler 2000]
Spiegler, I., Knowledge management: a new idea or a recycled concept?, in: Communications of the AIS, 3, 2000, No. 4, pp. 1-24
[Spiegler 2003]
Spiegler, I., Technology and knowledge: bridging a "generating" gap, in: Information & Management, 40, 2003, No. 6, pp. 533-539
[Tozer 1999]
Tozer, G., Metadata Management for Information Control and Business Success, Artech House, Boston 1999
[Tuomi 1999]
Tuomi, I., Data Is More Than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory, in: Journal of Management Information Systems, 16, 1999, No. 3, pp. 103-117
[Wang/Strong 1996]
Wang, R. Y., Strong, D. M., Beyond Accuracy: What Data Quality Means to Data Consumers, in: Journal of Management Information Systems, 12, 1996, No. 4, pp. 5-34
[White/Radcliffe 2008]
White, A., Radcliffe, J., Vendor Guide: Master Data Management, G00161285, Gartner Research, Stamford 2008

Appendix A Focus Group Participants

A. 1 Focus group interview on November 25, 2008

On November 25, 2008, 19 subject matter experts participated in a focus group interview in the course of a conference of the MDM Workgroup of the German-speaking SAP User Group (DSAG). In the 120-minute discussion the first version of the Functional Reference Architecture was assessed and recommendations for adaptation were given. The following table shows the participants.

Name	Company	Function
Jan Appl	Mieschke Hofmann & Partner	Head of Competence Center Strategy, Architecture & Methods
Jörn Bachmeier	Roche Diagnostics GmbH	Head of Global Material Master Management
Bernd Binder	Steria Mummert Consulting AG	Principal Consultant
Rainer Buck	Roche Diagnostics GmbH	Head of Global Material Master Management
Jens Edig	T-Systems Enterprise Services GmbH	Project Manager and Consultant
Gökhan Enç	SAP Deutschland AG & Co. KG	MDM Consultant
Helge Enenkel	Voith Paper Holding GmbH & Co. KG	Business Processes & Information Technology
Frank Fäth	ISO Software Systeme GmbH	Head of Division DQM Sales SAP
Bernd Gerhard	Deloitte Consulting GmbH	Senior Manager
Marc Koch	Koch BNC AG	Consulting Manager
Thomas Kübler	Adolf Würth GmbH & Co. KG	Head of Order Processing and Processes
Tim Merkle	IBSolution GmbH	Director SOA
Lars Metz	cbs Corporate Business Solutions Unternehmensberatung GmbH	Senior Project Manager Corporate Data Management
Johannes Mroz	ABeam Consulting (Europe) B.V.	Senior Manager
Manfred Nielsen	Karl Storz GmbH & Co. KG	Head of International Master Data Management
Martin Nussbaumer	IBSolution GmbH	Business Development Manager SOA
Hans Jakob Reuter	gicom GmbH	Managing Director
Wolfgang Sock	IMG AG	Consulting Manager
Karlheinz Sturm	Voith IT Solutions GmbH	Head of Master Data Management
Udo Zabel	aseaco AG	Managing Consultant

A. 2 Focus group interview on February 9, 2009

On February 9, 2009, five subject matter experts participated in a focus group interview in the course of the IRR Data Management Conference. In the 60-minute discussion an enhanced version of the Functional Reference Architecture was assessed and recommendations for adaptation were given. The following table shows the participants.

Name	Company	Function
Günther Engeler	Helsana Versicherungen AG	Head of Quality Management PK
Ralf Jäger	Client Vela GmbH	Partner
Detlef J. Königs	Mars Service GmbH	Business Data Manager Europe, Supply Chain Development
Norbert Schattner	Helsana Versicherungen AG	Solution Designer (technical/business) DWH
Jörg Stumpenhagen	Just.dot GmbH	Managing Director
Hans-Bernhard Wiesing	Corning Cable Systems GmbH & CO.KG	Global Data Management Organization Leader

A. 3 Focus group interview on February 18, 2009

On February 18, 2009, nine subject matter experts participated in a focus group interview in the course of a two-day workshop of the CC CDQ. In the 60-minute discussion a further enhanced version of the Functional Reference Architecture was assessed and recommendations for adaptation were given. The following table shows the participants.

Name	Company	Function
Cem Dedeoglu	Beiersdorf Shared Services GmbH	Head of Team BSS Master Data
Dr. Christian Ferchland	DB Netz AG	Strategic Infrastructure Data Management, Railway Geographical Data
Heiko Gebhardt	B. Braun Melsungen AG	Head of Central Material Master Agency
Dr. Federico Grillo	Beiersdorf AG	Head of Supply Chain Data Process Management
Gerhard Gripp	Bayer CropScience AG	Integration Manager Enterprise Master Data Managemet
Hans Jacoby	DB Netz AG	Head of Strategic Infrastructure Data Management, Railway Geographical Data
Jürgen Lay	Geberit International AG	Head of Group Product Data Management
Dr. Vlado Milosevic	ABB Information Systems Ltd.	Master Data Consultant and Headquarter IS Architect
Andrea Weissmann	Aesculap AG	SAP Inhouse Consultant Development and Master Data Management

Appendix B Functional Reference Architecture − Overview

A Master Data Lifecycle Management	**1** Data Creation	**2** Data Maintenance	**3** Data Deactivation	**4** Data Archiving
	Conditional Entries	Check-out	Bulk Editing	Archiving
	Bulk Editing	Bulk Editing		History Control
	Plausibility Check	Plausibility Check		
B Metadata Management and Master Data Modeling	**1** Data Modeling	**2** Model Analysis	**3** 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	
C Master Data Quality Management	**1** Data Analysis	**2** Data Enrichment	**3** 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	
D Master Data Integration	**1** Data Import	**2** Data Transformation	**3** 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	
E Cross Functions	**1** Automation	**2** Reports	**3** Search	**4** 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			
F Administration	**1** Data History Management	**2** User Management		
	Data Lineage	User Interface Design		
	Last User	Roles and Rights		

Figure B-1: Functional Reference Architecture – Overview