Electronic Government and Electronic Participation
E. Tambouris et al. (Eds.)

doi:10.3233/978-1-61499-570-8-124

Establishing a MDM function: First Steps in the Master Data Management Architecture Design

Riikka VILMINKO-HEIKKINEN[1]
Department of Information Management and Logistics, Tampere University of Technology, Finland

Abstract. Public sectors need to manage data for internal and external use is growing. Master data management aims to manage the core data that affects generally the data quality in large extent. The objective of the research is to observe the factors that affect the architectural decisions when establishing a MDM function. This is done though prior research and a case study in a municipality. Business needs and the existing IT environment indicate the best usage scenario for MDM. Scenario indicates the suitable architectural implementation style and also helps the organization to comprehend what they should emphasize in their development.

Keywords. Master data management, MDM, data management, Business intelligence, Open data

1. Introduction

Open data, big data and the growing need for predictive analytics set demands for public sector. Better data management practices are needed to ensure high quality of data for internal use, external use, and re-use. Master data management (MDM) aims to manage an organization's core data (i.e., master data). It tackles the data quality issues through process improvement by using organizational and technical aspects.

The focus in prior research has been in implementation styles and application architecture on private sector. Establishing MDM and the first steps of MDM architecture design have gained less attention. MDM development is a relevant topic in public sector in Finland. From September 2011, the Act on the Direction of Public IT Governance has mandated the use of an Enterprise Architecture (EA) Framework. The National Enterprise Architecture (NEA) steers the EA development [8]. Also MDM is a part of the NEA and the National MDM reference architecture was finalized in 2013. The objective of the research is to observe the factors that affect the architectural decisions when establishing a MDM function. This is done though prior research and a case study in a municipality.

The paper is organized as follows. First, we identify how prior literature has observed the architectural questions in MDM. The research methods and settings are then described. Empirical part follows and concludes with a discussion and recommendations for future research.

[1] Corresponding Author.

2. Related research

MDM should be seen as a process improvement plan to identify, assess, and implement methods and tools for assuring good quality data for the decision process [6]. The usage scenario defines the method of use for MDM.

In an operational scenario [10], [5], all interactions and transactions are applied to the master version, and consistency requirements must be strictly enforced [10]. Analytical scenario is relevant when the need is primarily to produce and maintain master data for data warehouse (DW), reporting purposes, analytics, and big data systems [10]. Enterprise scenario combines analytical and operational scenarios [1] and is similar to reference information management. The focus is on the importing of data into the master data environment and the ways that the data is enhanced and modified to support the dependent downstream applications. The collaborative usage scenario emphasizes achieving an agreement on a complex topic among a group of people [5]. This usually includes workflows and multiple tasks. Social MDM is a more recent perspective on MDM. It focuses on providing a platform for gathering, integrating, and stewarding a wider set of customer and product data, and for making them available through-out the organization [13]. Some of the unique characteristics of each usage scenario are presented in Table 1.

Table 1 Unique characteristic of different usage scenarios for MDM

Usage scenario	Characteristics
Operational	— Operational systems must execute their transactions against the master data environment instead of their own data systems [10], [5] — Integrates operational applications (e.g., ERP, CRM) in upstream data flow [15] — Individual and application access is closely monitored [10] — The MDM repository is considered as the authoritative source [11]
Analytical	— Applications are more likely to use than create master data [10], [14] — Applications can effect classifications/categorization of master data records [10] — MDM systems key role is to be a provider of consistent data for BI [5], [11], [10] — Intersection between the MDM and BI [5], mainly for reporting purposes [17] — MDM systems may include some key features for analytics [5] — Classifying master data is done through analytics [10] — Resembles customer data integration (CDI) [1] — Uses a unidirectional flow of data to the master record, using extract, transform and load (ETL) processes before importing the data [14] — Direct modifications to the master data can be made [10] — Least intrusive approach and most used [3]
Enterprise or reference information management	— Master records are created directly [10] — System coordinates users and systems to reach agreement on a data set [5] — Data in incorporated into the master environment, where it is available for publication to client applications [10] — Ensuring the quality of data in the entry [10] — Direct modifications to the master data can be made [10] — Classification made directly according to predefined rules [10]
Collaborative	— Collaborative authoring of master data: creation, definition, augmentation, and approval [8, 15]
Social MDM	— Analytics derived relationships [13] — Especially affiliated to big data capabilities [13] — Customer centricity (people and organizations) [13]

Repository, registry, and hybrid are common models for implementing MDM architecture. In repository, the complete collection of master data is stored in a single database

[10]. The global attributes of the data set are always created in central master data system [9]. In registry model, data sets are created, maintained, and distributed by different applications [9]. The hybrid model includes features of both of these approaches [10]. Also a consolidation model has been identified [16]. Fragments of master data are authored in a distributed fashion and stored in the source systems, but the central MDM system creates a composite golden record. Architectural models complement each other, and several models are sometimes used to fulfill the needs of an organization [18].

Master data application architecture provides complete overview of how the architecture is deployed by using different applications [2]. It contains applications for creating, storing and updating instances of the master data attributes defined by the conceptual master data model [19].

Many of the case studies addressing MDM architecture design (e.g. [3, 20, 21, 23]) have been focused on the private sector, the target of this research is a municipality, which offers an interesting opportunity to observe the architectural design decisions in a public sector organization. The focus will be on different MDM usage scenarios that have not received much attention in prior research.

3. Research Methods

An ethnographic case study (following the instructions of Myers [12]) was conducted in a municipality comprising 220,000 inhabitants and approximately 14,500 employees. The case organization consists of central administration, purchasing unit, welfare services, municipal corporations and several subsidiaries. The MDM projects and development were mainly conducted in the central administration.

The data collection period lasted from November 2010 to June 2013. During this period, two MDM development projects were carried out, and MDM development became rooted as a part of routine operations. The data was collected by participating in all project-related meetings and informal discussions in both projects, and in the project preparation and procurement phases of the second project. The author was actively involved in the first project as a member of both the steering group and the expert group. In the second project, the author acted as a project manager and as a member of the project group and the steering group. These positions offered unique opportunities to observe and thoroughly understand MDM implementation.

Diary entries were made weekly and whenever MDM-related issues were observed. In addition to observations, also questions that emerged and impressions were documented. To complement the diary, different kinds of project documentation were also utilized: procurement documentation, project plans, monthly status reports, and a set of memos from the working group, steering group, project portfolio group, stakeholder groups, and kick-off and closing seminars. Between the two projects, memos from the IT development group and the architecture group were also used. Finally, some internal documents were utilized, such as information management strategy.

The data was analyzed by following the interpretive research approach and the principles of Klein and Myers [7]. The data was re-viewed by observing it throughout and identifying what kind of objectives were presented during the process for establishing the MDM function. These objectives were categorized under several themes. Then these themes were used to review to data again. This time all issues supporting the themes were gathered. As a result, a list of the objectives was formed and discussion around these objectives was analyzed.

Ethnographers need to balance subjectivity and objectivity. Ethnographies are expected to meet standards of objectivity even when ethnographic research is highly dependent on the individual's unique knowledge and experience, and his/her actions as a thinking agent who brings his/her subjectivity to bear on the construction of information and knowledge [20]. All materials were analyzed in their entirety at the end of the overall data collection in June 2013. The idea was to gain some distance between the researcher and the context, and to keep data entries as neutral as possible so that they were not limited or affected by the analysis of earlier entries. This was done to minimize unintended entry manipulation, as one may easily make subconscious decisions about what to record.

4. Findings

The organization's centralized IT unit has been in operation since 2007. Previously, the units acquired information systems separately, with the exception of some organization-wide systems (e.g., Enterprise Resource Planning, ERP). As a result, it has approximately 400 information systems from different operating areas. The motivation for starting the MDM process was problems with data quality in main business processes. It was assumed that the problems originated from both maintenance processes and applications. The business objectives for MDM were identified for the first time in 2008. These were more efficient work, improved reporting, and service oriented architecture (SOA) interoperability.

The organization has several different external stakeholders (e.g. government agencies) that the organization is obligated to report regularly and also ad hoc (diary: 1/12). Data quality was considered one of the barriers to generating high-quality reports (BI report: 1.2.13). One of the clearest objectives became first supporting internal and external reporting (Diary 11/11) and later on supporting the BI comprehensively (Diary: 2/13).

MDM also had a role in harmonization of data structures, which would ease the difficulty in combining information (Diary 9/11). The last of the business objectives "service oriented architecture (SOA) interoperability" was not current later on in the development. The organization made the decision (EA principle) not to use SOA in the development of new applications in general.

The organization has been struggling with problems regarding data maintenance. Formal processes were inadequate and employees had invented additional ways to maintain the data to solve problems. One of the basic objectives is to simplify the process for data life-cycle management and to automate functions that had been performed manually. A large amount of the master data was still stored manually in Excel sheets, making maintenance difficult and error prone (diary: 3/12). For example, organization's products and services are managed manually (Diary 10/11).

Several steps should be followed when creating, changing, or deleting data (Business Workshop: 27.9.12). It is important that checkpoints for changes are in place and those checkpoints are automated (Diary: 1/12). Workflows are particularly important when there are several tasks in the workflow or when a task is performed less frequently (Diary: 4/12). The data quality validation should be done while creating the data. History of the data should be also stored. Data standards were seen as an important issue in tackling data quality issues (Steering Group memo: 11.10.11). Several roles

were identified to enable a finely divided control of the data. Roles were strictly limited to the need to make changes (Security and Privacy Work-shop: 13.9.12).

The organization's master data included sensitive attributes. The ongoing situation was that this information is often in several applications and there were problems with access management. Data is also imported from external sources. Usually this was done separately to the individual applications and the data would be obsolete in this respect in other applications (Diary 4/12). MDM should support compliance and provide a reliable foundation to support changes and updates in policies to help avoid penalties or other regulatory actions (Diary 8/12). The MDM system´s log should make it possible to identify problems relating to data misuse or other issues. For example, there are examples about data of organizational unit being deleted incorrectly (Diary: 9/11).

Organization´s master data domains include several hierarchies. Maintenance of these should be coherent (Diary: 1/12). The hierarchies should match those used at government level, and the maintenance should be synchronized (Project Group memo: 30.1.12). Master data is also affected by the definitions of government and government agencies and the need for a change often comes from a stakeholder. These affect the modification needs of the data models, attributes, and also the metadata (Business Workshop: 27.9.12). Master data objects should be categorized in different ways to present the perspectives that the data is used and observed. This serves especially the needs of the BI.

Several ways for enhancing data quality were identified (IT Workshop: 28.8.12, 4.9.12, 7.9.12, and 18.10.12). External sources should also be used for validating the data (IT Workshop: 28.8.12). The need for methods to continuously monitor the data quality was also identified (Project Group memo 28.9.2011; Diary: 4/12; Diary: 9/12).

The organization outlined an initiative to open data in a machine-readable format as part of their operations (Diary 5/12). Master data has high value in terms of re-use (also for commercial use) (Diary: 10/11) and in making the organization's operations transparent. It was seen as a problem that the data was scattered. For this purpose the structure of the data as well as easy access was essential. Master data often includes attributes that cannot be opened, it is important that there is attribute level access control that helps to control what data is published and by whom. Also, the understanding in the organization about this is part of the data governance and data privacy (Diary: 4/12; Security and Privacy Workshop: 13.9.12).

5. Discussion

The case organization's business needs were observed and classified. These were categorized under different usage scenarios. Summary is presented in Table 2.

Table 2 Business needs categorized by usage scenarios

Reference/Enterprise	Analytical	Operational	Collaborative
Control and formalize data creation	Automate infrequent tasks	Streamline work processes and the organization	Workflows, approval points
Function-oriented instead of system-oriented doing	Generate automatic reports of changes for stakeholders	Control and formalize the creation of data, Reduce errors in data entry	Decentralized data creation
Reduce errors in data entry	Automated and secure publishing	Access control to sensitive data	Divided control of the data
Direct data entry through one UI	Publishing machine-readable open data	Comply with security and data privacy rules	

Easier master data quality management and enhancement	High quality data source for DW (relationships between the concepts)	Better interoperability and view of the organization's core data
Supporting data maintenance	Data classification	High-quality data for process-es
Data standards, classi-fication based on predefines rules	High-quality data for reporting and enabler for BI	Easier master data quality management and enhance-ment
Modification of metadata		Improve the data quality
Data categorization		Use of data standards
Detect errors before they affect functions		Approval function and formal process for maintenance
Comply with security and data privacy		Supporting the maintenance of data

Many of the objectives responded to analytical scenario's characteristics. Despite this; it was also quickly evident that the scope was wider. Reporting was seen as an important area, but the main problems affecting it were issues with data quality in operational systems. The development of BI was ongoing and the clarifications in the BI development also shed some light to MDM development. Big data development was observed in the last phase of the data collection. The analytical scenario was not efficient for the organization in long run. As seen in Table 2, business needs support almost all of the scenarios in some extent. Collaborative scenario seemed very separate from the other scenarios. It seems to be more a perspective on the other scenarios than a distinct method of use. Social MDM approach would acquire a more mature phase of big data development and use. The vision for BI included, e.g., idea to use social media data to enrich the customer data, but it there was no clear development plan for big data.

Number of the objectives supporting the scenario was not adequate to refer to the suitable scenario. Cervo and Allen [3] emphasize indirect objectives, such as cost savings, as assessment criteria, but in the case organization these were not considered at this stage. Cost savings were much more seen as a result of the business needs, instead of a separate factor. Because of this, organization and IT-landscape were also observed as factors that emphasize the suitable scenario option.

The model of centralized IT would support the operational scenario. Organization's IT environment is complex and for the legacy systems, publication to other applications as defined in reference architecture would be more interoperable with these. Operational usage scenario seemed to echo to the business needs and IT environment most. Because of this and the needs for analytical usage, the organization's usage scenario was enterprise MDM.

The MDM usage scenario narrows the options for architectural implementation styles. Certain styles respond better to certain usage scenarios [3]. At the final stage of the data collection, the organization decided the hybrid model as an implementation style. This style is suitable when the organization is looking for a method to improve and manage the data quality, completeness and consistency of master data across several systems and the organizational commitment and the re-sources support proper data governance activities [16]. Dreibelbis et al. [5] have also identified the characteristics of legacy systems as constrains for choosing an architecture pattern for MDM. The organization's ERP was seen as a master system for two of the identified master data domains. Hybrid model has been identified as a good fit for ERP environments [11].

The current IT-landscape was the main influencer for making the decision of the MDM application architecture. The large number of legacy systems was one of the reasons why MDM architecture included a separate MDM system. The organization struggled to make the decision between a dedicated MDM system and an operational system converted to a master data source. The amount of the existing applications was the reason why they were reluctant to acquire a new system. Also the cost of a new solution was seen problematic. In the end, the extent of the desired MDM development resulted in the need for MDM system to enable the required elements and functions.

Prior research has not addressed MDM in the public sector or especially in the municipalities, nor has it made generalizations about the common or distinguishing features between the public sector's and private sector's MDM. The type of the organization affected the MDM development in many ways, but did not seem to affect significantly the first choices made with the architectural design. However, the MDM application architecture was affected by the existing IT-landscape.

6. Conclusions

The objective of the research was to observe the factors that affect especially the first architectural decisions when establishing a MDM function. The prior research names three layers in the MDM architecture design. First step includes identifying the method of use for MDM. Second step is to determine the right architectural implementation style for MDM. Last step, the application architecture, defines the technical architecture in detail. The first step has gained less attention and because of this, it was observed more closely through the case study.

The factors affecting the first architectural decisions were complex. The maturity stage of the organizations BI responded well to analytical usage scenario. For an organization that has higher maturity, social MDM could respond well to the needs. Operational view contemplated the demands for streamlining work processes, and enhancing data quality in operational systems and processes across the organization, and the needs that open data places. It also responded well to the data security and privacy demands. Because there were characteristics of multiple of the approaches, the usage scenario could be identified as enterprise MDM. Business needs should indicate primarily the chosen architectural scenario, but also other factors have an effect. IT environment was the final factor in identifying the suitable usage scenario. This also indicates that the architectural implementation style should respond to different use scenarios.

The MDM architectural design is complex and challenging to design. To start the process by identifying the usage scenario for MDM through business objectives and IT environment sets the architectural path to right direction. Usage scenario indicates the suitable architectural implementation style and also helps the organization to comprehend what they should emphasize in their development. MDM usage scenarios have been dismissed in prior research for not being elaborate (e.g. [14]), but it seems that they present a good starting point for designing MDM architecture. Usage scenarios imply the applicable architectural styles and further on also the suitable application architecture. They also clarify how objectives are translated into design decisions for the organization.

Public sector's master data management practices have gained very little attention in research. Still, these should be emphasized more, because their data is transforming into public data. Public sector organizations set interesting research settings because of

their complex IT landscapes and diverse data domains. MDM has some unique characteristics when established in public sector (e.g. data privacy issues). This research was a single case study, and caution should be exercised with regard to generalizations. For future research, there are several different paths to explore. The effects of usage scenarios to the design of governance might offer interesting perspectives on the overall concept of MDM. The social MDM in big data development also offers an interesting viewpoint.

Acknowledgements

The author wishes to thank Professor Samuli Pekkola who provided insight and expertise that assisted the research.

References

[1] C.G. Apostol, Enterprise Master Data Management Trends and Solutions. *Informatica Economica*, 3(43) (2007), 35-39.
[2] E. Baghi, S. Schlosser, V. Ebner, B. Otto, H. Österle, Toward a Decision Model for Master Data Application Architecture. *In: 47th Hawaii International Conference on System Science*, (2014), 3827 – 3836.
[3] D. Cervo, M. Allen, *Master Data Management in Practise: Achieving True Customer MDM.* Hoboken, New Jersey: John Wiley and Sons, 2011.
[4] T.P. Das, R.M. Manas, A Study on Challenges and Opportunities in Master Data Management. *Int. J. of Database Management Systems*, **3**(2) (2011), 129-139.
[5] A. Dreibelbis, E. Hechler, I. Milman, M. Oberhofer, P. van Run, D. Wolfson, *Enterprise Master Data Management: an SOA Approach to Managing Core Information.* Westford, IBM Press, 2008.
[6] M. Fung-A-Fat, Why is Consistency So Inconsistent? The Problem of Master Data Management. *Cutter IT journal*, 20(9) (2007), 23-29.
[7] H.K. Klein, M.D. Myers, A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. *MIS Quarterly*, **23**(1) (1999), 67-93.
[8] J. Lemmetti, S. Pekkola, Enterprise Architecture in Public ICT Procurement in Finland. *Electronic Government and Electronic Participation* (2013), 227 – 236.
[9] C. Loser, C. Legner, D. Gizanis, Master Data Management for Collaborative Service Processes. *In: International Conference on Service Systems and Service Management* (2004), 1-6.
[10] D. Loshin, *Master Data Management*. Burlington, MA, Morgan Kauffman, 2008
[11] A. Maedche, An ERP-centric Master Data Management Approach. *In: 16th Americas Conference on Information Systems*, Paper 384 (2010).
[12] M.D. Myers, Investigating Information Systems with Ethnographic Research. *Communications of AIS*, **2** (1999), 1-20.
[13] M. Oberhofer, E. Hechler, I. Milman, S. Schumacher, D. Wolfson, *Beyond Big Data: Using Social MDM to Drive Deep Customer Insights*. Pearson, Indiana, 2014.
[14] B. Otto, How to Design the Master Data Architecture: Findings from a Case Study at Bosch. *Int. J. of Information Management*, **32**(4) (2012), 337-346.
[15] R. Silvola, O. Jääskeläinen, H. Kropsu-Vehkaperä, H. Haapasalo, Managing One Master Data - Challenges and Preconditions. *Industrial Management & Data Systems,* **111**(1) (2011), 146-162.
[16] A. White, D. Newman, D. Logan, J. Radcliffe, *Mastering Master Data Management*. Stamford, Gartner Research, 2006.
[17] A. White, J. Radcliffe, *Four dimensions of MDM*: Understanding the complexity. Stamford, CT: Gartner research, 2007.
[18] R. Vilminko-Heikkinen, S. Pekkola, Establishing an organization's master data management function: a stepwise approach. *In: 46th Hawaii International Conference on System Sciences* (2013) 4719-4728.
[19] R. Winter, R. Fischer, Essential Layers, Artifacts, and Dependencies of Enterprise Architecture, *J. of Enterprise Architecture*, **3**(2) (2007), 7-18.
[20] U. Schultze, A Confessional Account of an Ethnography about Knowledge Work, *MIS Quarterly*, **24**(1) (2000), 3-41.