Paper ID #14390

Architectural Evaluation of Master Data Management (MDM): Literature Review

Dr. Gholam Ali Shaykhian, NASA

Ali has received a BS degree in Electronics, a Master of Science (M.S.) degree in Computer Systems and a second M.S. degree in Operations Research from the University of Central Florida and has received a Ph.D. degree in Operations Research from the Florida Institute of Technology (FIT). His research interests include knowledge management, data mining, object-oriented methodologies, design patterns, software safety, genetic and optimization algorithms. Dr. Shaykhian is a professional member of the American Society for Engineering Education (ASEE), serving as the past Program Chair for the Minorities in Engineering Division. He has served as a chair, vice-chair, program chair, and program committee member for numerous conferences of ASEE.

Dr. Mohd Abdelgadir Khairi, Najran University

I, Mohamed Khairi, my bachelor degree in computer science. I did my Masters in system science from University of Ottawa, Canada. My PH.D was in "Master Data Management" from University of Phoenix. I have over 20 years of experience in IT industry - ten of them with Microsoft in Redmond, WA. Currently I'm assistant professor at University of Najran. In addition of teaching and Research I'm coordinator of graduation projects and field training for computer and information system college.

Dr. Jinan Ziade,

I have a PhD in Organizational Leadership with emphasis in IST from University of Phoenix, and an MBA from the same university. I have over 7 years of extensive leadership experience in advertising, marketing, strategies, and project team lead. Currently serving as Program Chair of Guild volunteer at St. Jude Medical Center and working with Memorial Foundation on philanthropic endeavors. My research interest include knowledge of cultural differences and leadership practices within global organizations and problem solving. Developed a leadership practices model that provides a suitable framework and salient business strategy component for corporations seeking to expand successfully.

Architectural evaluation of Master Data Management (MDM): Literature Review

Abstract: Architectural evaluation of MDM data models are influenced by technical assessment, business requirements, and designers' preferences. Therefore, selecting the right model for an organization requires quantitative method with qualitative embedded design study. Greene [4] states that three purposes of quantitative study with embedded qualitative design include 1) triangulation, to guarantee or attain validation of data, or convergent validation; 2) complimentarily, to elucidate, clarify, or otherwise more fully detailed the results of analyses; and 3) development, to lead the use of additional sampling data gathering, and analysis procedures (p. 259). The existing literature demonstrates the need for both qualitative and quantitative design analysis to describe the relationship between MDM architectural models' alternatives (federated and centralized).

Current literature studies do not dispute the importance of MDM models and their relationships to the single view of organization's master data. The IT community unanimously supports the notation of "single view of organization's master data." According to Rosenberg [16], master data forms the basis for the business process; Loshin[9] claims that the main goal of designing any MDM is to ensure that the system is business driven. Loshin [9] considers the implementation MDM a failure if it fails to address the business needs (p. 54). This paper documents the importance of literature study to support the need for qualitative and quantitative analysis.

Introduction: "Master Data Management (MDM) can be defined as an enterprise-wide infrastructure for integration, harmonizing and managing of master data, so that information is supportive of business decisions to enhance organization value" [1], P. 1). Master data consist of a multiple entities. For example, master data may consist of customers, products, locations, and calendars [8]. MDM maintains an organization's master data and organizations benefit from MDM services through Service Oriented Architecture (SOA) [10]. The major domains that use MDM data management have the capability of maintaining a single reference point for master data are product information management, customer data integration, and vendor information management [17]. In addition, the major objective of MDM system is to support and manage consistent master data of all operational systems [10].

For many years, architectural models have been used by electrical, mechanical, and structural engineers. Computer science and information systems, compared with other engineering disciplines are relatively new and hence do not adhere to the same rigorous set of rules. Architectural framework in engineering disciplines has proven to be of significant values; today, no one builds a building without going to multiple layers of architectural review and analysis. The literature study is motivated by these very facts and is aimed at evaluating the existing IT architectural frameworks to enhance the state of research and contribute to IT field.

MDM's architectures alternatives: The advancement of databases technologies have resulted in proliferations of data, databases, and data management. MDM was introduced as a way to improve efficiency of master data access and management. In the MDM model, federated and registry have the same meaning whereas centralized means repository or transactional. The terms architecture, model, and style was used in the study. In addition, the term hub is used for MDM data location. Wolter and Haselden [18] suggested three basic types of MDM models:

1) Registry (Federated) – Registry model contains only the primary key and the name of the source system. The master data remains in the source system. The registry case requires mapping to source system. Dreibelbis [3] suggested that Registry model is a central reference and is read only hub and the master data creation, the performance of data modification or addition can only be in the source systems.
2) Repository (centralized), in which the master data is moved to MDM and the application systems consumes the master data from the MDM repository hub. Unlike the Registry model, the mapping to source system is not a necessity. Many authors labeled Repository style "Transactional" because it is a "system of record" (SOR) in which master data creation, addition, deletion, and modification done in MDM hub [3].
3) Hybrid (CDM and FDM combination), which is the mix of both Registry and Repository. Dreibelbis [3] used the "Coexistence hub" term for mixed style because it is main purpose is to harmonize the sources with MDM hub and master data exists in both sources and MDM hub. Hybrid style is not system of record and is read only.

Federated (aka Registry) MDM Model: Organization's data may be categorized as its enterprise data when is available to different application databases. Enterprise data is considered federated or registry when it is distributed in different locations and available for enterprise data integration through the services offered by the application. The entire enterprise may inquire the available and persistent master data through federation services in real-time or near real-time mode [10]. However, in federation data service, MDM acts as a hub whereas applications using the master data act as a spokes [10] (The term hub and spokes came from the airport analogy, in which a major airport acts as feeder (hub) for local airports (spokes)). Hence MDM is a feeder whereas master data sources are spokes.

The term "federation" is associated with Registry style MDM hub implementation. The Registry system based on federation model consolidates required master data together for business use from different sources [12]. The Registry style is attractive because 1) it could be easily and quickly implemented without facing political issues surrounding common data definitions and decisions agreement, 2) Because data is left in the source system, the data quality requirement will be minimal, 3) fast to build, and 4) the complete view is assembled as needed.

The biggest disadvantage of registry style is the inability to clean the master data despite its usefulness in finding duplicate and inconsistent records because the data remains in the source. For example, if customer A is in seven records in the CRM database, four records the ERP database, and three records in the customer support database, and among the 14 records there are six phone numbers, a registry hub will not help to get CRM and EPR systems to agree on a phone number; however, the registry will tell the user the location of all records [18]. The other disadvantages of FDM are that it is a read-only system and it is not always current with

other data sources. The main advantages of a registry model are 1) MDM is lightweight because it has only key data, 2) easy and fast deployment because the moving of amount of data is easy, 3) is scalable and is the least intrusive, and 4) easier to update because it is easy to add or change to attributes or interrelationships between master data and the source [12].

Madhukar [10] believes the data federation is a suitable option for companies that need a quick solution, but want to minimize the risk of deploying a new initiative. It may help organizations that do not have the budget to support staffing, and infrastructure centralized model needs. For organizations with a high volume of data, federation will be suitable option to solve performance issue. Federated data model and services enable an organization to create views from multiple systems without draining organization's system's resources. The maintenance of a federated system is easy. Maintenance will not cause major failed customer interactions. For instance, evening maintenance may occur while other components are active [10].

FDM has may benefits for MDM applications customers: 1) single version of the customer data is accessible, 2) highly reliable and scalable SOR system that serves all the LOBs (Line Of Businesses) and Channels in the organization, 3) increased sales revenues because of more cross sell and up sell Opportunities, 4) lower TCO (Total Cost of Ownership) because of lesser maintenance costs as the existing SOR s will slowly retire, 5) increased customer satisfaction and retention, 6) consolidated data enables rapid sales and marketing turnaround times, 7) fast performing SOA web services resulting in increased productivity, 8) analyze the consumer pattern analysis through intelligent analytics, 9) ability to monitor the application health through intelligent and real-time monitoring dashboards, and 10) enhanced application maintenance and scalable, reliable catering to future needs [10]. The other advantage of FDM over CDM is the synchronization between the application sources and MDM hub is not expensive. In the case of CDM, reengineering of source applications is required whereas with FDM it is not necessary. The performance of source system has an impact on the MDM hub that architected in a registry model and it is the main drawback of a registry model because MDM performance depends solely on domains performance [12]. The master data in FDM may not reflect accurate master data in case of source system in state of maintenance or failure.

Madhukar [10] addressed five challenges in implementing federated data domains 1) continuous synchronization of the Hub and the federated domains because the requirement for correct view of data remains high, 2) SLAs (Service Level Agreements) for online and near real-time processes are critical for the success of the MDM to derive best data quality, 3) one-time data load may cause data quality issues, 4) maintenance of the legacy SOR (System of Records) systems until the acceptance of federated MDM solution as the Single SOR for the organization, and 5) change Management and training.

Centralized (aka repository) MDM Model: The Centralized MDM, Repository, and Transactional hub are all terms that have the same definition. Authors such as Loshin [9] and Moseley [12] use the term repository instead of centralized. Dreibelbis [3] and other authors use the term Transactional hub. For the purposes of this study, the term centralized was be used. In

the Centralized model, all domains of master data are consolidated in one repository, manage a single view of master data, and provide data access through services. In contrast with the federated model, the centralized model resolves the issue of latency or synchronization because there is no update issue of copying the same master data more than once.

In a centralized model, a single copy of master data is easier to manage. Compared to registry model, the quality of data will be much higher than registry model and its issue of mapping and synchronization [18]. The centralized model is a system of record because it supports existing or new transactional applications [3]. On the other hand, there are many obstacles facing repository style implementations:

1) Politically and technically difficult to identify and decide common data for all applications within the organization.
2) Transform and load the data from current databases into the MDM hub while cleaning the data in the process.
3) Reengineering all current application to obtain master data from MDM hub; however, reengineering could be an enormous effort.
4) Handling the historical data records. Wolter and Haselden [18] identified the cost and difficulty to overcome the above four obstacles to justify ROI (Return On Investment). He added Repository style may not be suitable for many projects.

Even so, he didn't elaborate more on whether or not the implementation of repository will provide better ROI in the future. Another drawback of CDM is that it may request reengineering the source systems to adjust to access and retrieve master data from external central repository. This change may include security changes, transactional and operational changes, expiration of the old system of managing master data, and change the interface. The advantage of CDM over FDM is that FDM is analytical only while CDM is collaborative, operational, and analytical in the method of use.

According to Dribelbis [3], the CDM has the following advantages and characteristics: 1) make MDM as a system of record, 2) uses the method of analysis, 3) collaboration, and operational, 3) MDM serves as single source of master data, 4) master data created and maintained only in MDM system, 5) allow correctness of master data, and 6) master data is complete and consistent. Moseley [12] believes that the ability to make change in one system instead of multiple systems is the advantage of a centralized model. The centralized model keeps the "system of record." The primary disadvantage of centralized model is that it is not supported by most commercial off-the-shelf (COTS) source systems such as ERP because they are independent and self-contained. Therefore, the source systems are difficult to change according to centralized model's requirements. The second disadvantage of the centralized model is that the performance may be affected by the demands of all transactions that need master data. The third disadvantage is the difficulty of making centralized systems highly available because of possible maintenance that takes the system offline. Finally, the ROI of a centralized model is not immediately apparent and it may be difficult to sell the idea to business [12].

The challenge with centralized model is that isn't suitable for large organization to implement the model because it is specifically suitable for organization with low transaction and less data volume [12]. For that reason, Moseley [12] discourages large companies from implementing centralized MDM solution. Dribelbis [3] believe moving master data from an unmanaged state in initial load to managed state is a challenge for centralized model implementation. However, Dribelbis [3] suggested moving the master data to repository requires Extract Transform Load (ETL) process compose of cleaning and de-duplicate of mater data during the extract and transform; and suggested the federated model as an alternative to the difficult and expensive ETL process.

MDM Architectures Focuses: Technology (MDM products), process (best practices to manage master data) and people (the requisite knowledge and skills to design and implement MDM) are the main focus of MDM strategy. Dreibelbis [3] described the MDM strategy as getting the value of combining technology and business. The author insisted on making sure the business motivation is sufficient to support the MDM strategy that includes business guidance and participation through the process.

MDM technology implementation should be scalable, able to adjust to business requirements changes over time. Architectures' alternative choices should consider organization needs and ability. MDM products and technologies are essential ingredients for a sustainable solution. Organizations must use best practices that address the MDM end-to-end scope. The best practices include data governance, system interfaces and integration, application build, system deployment, and operational aspect of MDM (p. 303). The MDM architectures focus on three areas (which we will discuss in detail below): Analyzing how the system environment react to business uses, technology focused, and business focused approach questions.

Business focused approach questions: To choose the best MDM architecture, whether the model is FDM, CDM, or hybrid for an organization, the two categories of in-depth analysis must be done. The two analyses are: 1) Business case analysis, 2) system environment use metrics analysis. A business case analysis is the process of understanding the use of the system to create the value of improving specific processes that company needs. An analysis of system environment use metric, which will make sure the support of availability, scalability, and reliability requirements are met. The in-depth analysis will trigger an architect to seek answers for the following three questions:

First: Tackling business problem question. Mainly, business has two major problems with the master data; consistency and identification. Having consistent master data that will enable an organization to: 1) save money, 2) increase customer satisfaction [6], 3) increase productivity, and 4) improve reporting [7]. On the contrary, inconsistent master data cause process errors and thus higher costs [2].

Second: The business use question. MDM initiative is more business focus than other IT discipline. MDM implementation is mainly to serve and support business goals. Unlike BI,

MDM has a significant operational use. MDM design and implementation should adhere to business needs and requirements. In other words, there could be many different MDM implementations depending on business objectives [1]. Individuals who work in MDM space must build a business case to justify the needs to sell it. Power [14] stressed the importance of building business case for MDM application for other people in the organization that do not work in MDM but strive to understand the cost, the value, and MDM implementation benefits.

Third: The business requirements for master data governance question. MDM attracted companies that want to save their data and view it [15]. MDM is an effective approach to address the data quality issue. Data quality requires analysis; technology, and business involvement, which means strong data governance. MDM could help identifying the data quality problems, clean the data, and synchronize the data between systems [11].

Technology Focused: Gruman [5] warned against ignoring the data architecture and insisted on the need for application development and structure model.

Commercial Off- the- shelf (COTS) solution. O'Sullivan [13] provided the following software and solutions for MDM in the market: D&B/Purisma Data Hub; DataFlux MDM; Data Foundations OneData; IBM InfoSphere MDM Server; Kalido MDM; Microsoft MDS; Oracle Customer Data Hub; Oracle Hyperion DRM; SAP NetWeaver MDM; Siperian MDM Hub; TIBCO CIM; and VisionWare MultiVue. Before choosing off the shelf solution product, there are some criteria that need prioritization and availability for business needs. The following criteria are main criteria that organizations need to consider before MDM implementation. O'Sullivan [13] asserts that organizations can validate products according to the following criteria: 1) Support for Multiple-Entity types such as customers, accounts, products, and vendors; 2) User types and usage scenarios such as analysis and reporting; 3) Data governance and administration features and capabilities; 4) Support for unstructured data and search capabilities for it at enterprise level; 5) Support Centralized or decentralized repositories; 6) Ease to deploy; and 7) Pre-built Domain Data Model.

Security and privacy considerations: Security and privacy are crucial in MDM architecture and one must take security considerations into account in MDM implementation. Dreibelbis [3] discussed the considerations of security related in MDM implementation: First, Policy is part of general data governance plan. The privacy and security policies are the rules and process organization must follow to protect its assets and customers privacy. The security rules in MDM include the access level and authorization to view or modify master data.

Most of the commercial off-the-shelf (COTS) MDM software such as Microsoft MDS solution allows organization to provide a different level of access and authorization to master data users. Secondly, confidentiality consideration is important because master data is sensitive, the data and information must be disclosed only to individuals who are authorized to access and use it. Third, integrity that means unauthorized user can't modify data without detection. The

assurance of integration is necessary for privacy as well as security. Fourth, the process of validating and verifying the system or individual identity is called authorization. The system must validate the person or the system claim of authority to use data by comparing user's credential against policy and rules data. Fifth, MDM security policy identifies the users' authorization role to do specific activities with the data such as view or modify. Sixth, system activities auditing is necessary to track use and access and to understand system behavior regarding the security and privacy policies implementation. Seventh, users, systems, roles, groups, and other information about users must be registered and recorded in the system to define the authorization and authentication. Finally, reverse proxy, which is to verify the user requests are properly audited, authorized, and authenticated.

Summary: The major role that MDM plays in organization is to assist with attaining a single view of master data. The discussion in this study involved the scholars and expertise ideas in terms of MDM. The discussions included the best approach to attain the right decision for implementing MDM model to arrive at "single version of the truth." As Moseley [12] suggested the focus of an architect should include both business and technology analysis to select the suitable model for an organization. Many products are off-the-shelf available to solve the master data management. Nevertheless, one must carefully check the list of organization requirements against what the product may offer. The literature review in paper addresses the advantages, disadvantages, and challenges facing each MDM models and how to address the challenges to overcome the obstacles. These models when deployed can help to improve efficiency of an organization. Many authors have discussed the federation data model and how it can be deployed in enterprise architecture and the benefit of fast deployment and the data view assembled per request.

In addition, there are numerous literatures that address the centralized model and discussed CDM characteristics, benefits, drawbacks, and method of use. Many authors such as Loshin [9] and Moseley [12] used the term registry in association with federation, and the term repository for centralization. However, Dreiblibes et al. used the transactional model in association with centralization. Upon the point of completion, new applications can start using the MDM hub directly for their master data so the hub evolves gradually toward the Repository style. Although many organizations may not be able to move completely to the Repository style, eventually it may become the predominant approach for applications as the old applications are replaced. Hence, Wolter and Haselden [18] believe the registry is the starting point to reach the repository. Moseley [12] agreed with Wolter and Haselden [18] on certain points whereas Madhukar [10] asserts that using a repository model implementation for large organization is not the best model.

References

1. BIPM Institute (2008). Master Data Management definition- What is MDM-CDI?. Retrieved from http://www.bipminstitute.com/master-data-management/MDM-concept.php

2. Bijesse, J., Higgs, L., McCluskey, M.(2002). Service Lifecycle Management (Part 1): The Approaches and Technologies to Build Sustainable Competitive Advantage, Atlanta: AMR Research.

3. Dreibelbis, A., Hechler, E., Milman, I., Oberhofer, M., van Run, P., & Wolfson, D. (2008). Introduction to enterprise master data management. Retrieved from http://searchdatamanagement.techtarget.com/generic/0,295582,sid91_gci1320762,00.html

4. Greene, J.C., Caracelli, V.J., & Graham, W.F. (1989). Toward a conceptual framework for mixed-method evaluation designs. Educational Evaluation and Policy Analysis, 11, 255–274.

5. Gruman, G. (2006). Rolling out the real time data mart. *Infoworld*. Advance online publication.

6. Joshi, M., Subrahmanya, V. (2003). *Global Data Synchronization*, Bangalore: Wipro Technologies.

7. Kelly, R., Fritsch, M.(2002). Improved effectiveness with information consolidation. Retrieved from http://download-west.oracle.com/owparis_2003/40266.doc.

8. Kimball, R. (2008). Essential Steps for the Integrated Enterprise Data Warehouse, Part 1. *swtuopproxy.museglobal.com*. Advance online publication. Retrieved from http://swtuopproxy.museglobal.com/MuseSessionID=fd925b9615e67115f7e6173a6599d7e2/MuseHost=proquest.umi.com/MusePath/pqdweb?index=0&did=1454942261&SrchMode=2&sid=1&Fmt=3&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1258783964&clientId=13118

9. Loshin, D. (2008). Master data management. San Francisco, CA: Morgan Kaufmann.

10. Madhukar, N. (2009, June 24). Federated MDM data domains - A Perspective. Retrieved from http://www.infosysblogs.com/customer-relationship-management/2009/06/federated_mdm_data_domains_a_p_1.html

11. McKnight, W. (2009). Four more sources of poor data quality: How to develop a strategy to identify and understand the issues. *ProQuest*. Advance online publication.

12. Moseley, M. (2009, October 15). Part 7: Centralized Models: Complete but Expensive. [Web log message]. Retrieved from http://blog.initiate.com/index.php/2009/10/15/part-7-centralized-models-complete-but-expensive/.

13. O'Sullivan, T. (208). Multi-Domain and Federated MDM Hubs: Two Key Considerations. *blog.hubdesigns.com*. Advance online publication. Retrieved from http://blog.hubdesigns.com/2008/03/04/multi-domain-and-federated-mdm-hubs-two-key-considerations/

14. Power, D. (2009). The Best Pitch for Customer Data Integration; Follow these 10 steps to deliver a business case within your company's project requirements. *Information Management*.

15. Premisberger, C. (n.d.,). Change comes to MDM space. *Information Management*.

16. Rosenberg, J. M. (1987). Dictionary of Computers. Information Processing & Telecommunications, 2nd ed., New York: John Wiley & Sons.

17. White, C. (March, 2009). Can enterprise data warehousing and master data management projects survive the recession? Retrieved from HTTP://WWW.B-EYE-NETWORK.COM/VIEW/10010

18. Wolter, R., & Haselden, K. (2008). The What, Why, and How of Master Data Management. *msdn.microsoft.com*. Advance online publication. Retrieved from msdn.microsoft.com/en-us/library/bb190163.aspx